Journal of Agricultural Science; Vol. 7, No. 8; 2015
ISSN 1916-9752 E-ISSN 1916-9760
Published by Canadian Center of Science and Education

DNA Sequence Characteristics and Phylogenetics of Putative Imprinted Genes on Bovine Chromosome 29

Oladeji Bamidele[1,2], O. G. Omitogun[1] & I. G. Imumorin[2]

[1] Biotechnology Laboratory, Department of Animal Sciences, Obafemi Awolowo University, Ile-Ife, Nigeria

[2] Animal Genetics and Genomics Laboratory, International Programs, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA

Correspondence: Oladeji Bamidele, Biotechnology Laboratory, Department of Animal Sciences, Obafemi Awolowo University, Ile-Ife, Nigeria. Tel: 234-705-370-5484. E-mail: bamideledeji@gmail.com

Received: March 16, 2015 Accepted: May 13, 2015 Online Published: July 15, 2015

doi:10.5539/jas.v7n8p131 URL: http://dx.doi.org/10.5539/jas.v7n8p131

Abstract

Cattle are important livestock species with huge genetic resource for food security, agriculture and livelihoods. Over 60% of its genes are homologous to all mammalian species which creates a molecular basis for conducting comparative genomic analysis. Genomic imprinting has been implicated in a variety of biological functions and so identification of new or verification of known imprinted genes in livestock species is of high agricultural and biomedical importance. Fourteen (14) putative imprinted genes on bovine chromosome 29 (Bta 29) as well as the human (Hg 11) and mouse (Mm 7) orthologs were computationally characterized with respect to the CpG islands (CGI), transcription factor binding elements and sequence motif. Phylogenetic analysis was conducted across the three species for each of the genes identified to have promoter CGI. Promoter CGI were identified in *ASCL2*, *TSSC4*, *CDKN1C*, *KCNQ1*, *PHLDA2* and *NAP1L4*. The promoter CGI were enriched with CpG containing transcription factor binding sites. Generally, it was observed that cattle was more closely related to human than mouse and that natural selection was the force driving the evolutionary change between the three species. Protein kinase motifs involved in phosphorylation were identified in the amino-acid sequences of *ASCL2*, *TSSC4*, *PHLDA2* and *NAP1L4*. Our results suggest the post-translation regulation of imprinting and that the predicted promoter CGI can be assayed to determine molecular function, gene expression and DNA methylation status of the bovine putative imprinted genes.

Keywords: cattle, CpG islands, imprinting, phosphorylation, phylogenetics

1. Introduction

Genomic imprinting is widely studied due to its involvement in various biological processes such as brain function and behaviour (Garfield et al., 2011), tumorigenesis (Lim & Maher, 2010), control of intra-uterine growth and birth weight (Schulz et al., 2010), reprogramming during embryo and nuclear transfer as well as in somatic cell cloning (Kedia-Mokashi et al., 2011). It provides a mechanism to distinguish between the paternal and maternal genomes and also regulate biological processes. The number of experimentally confirmed imprinted genes is over 100 in mammals and studies indicate that up to 1,300 (Gregg et al., 2010) and as many as 2,000 additional genes (Nikaido et al., 2003) could still be imprinted in mammals. Most of these imprinted genes were identified in human and mouse. Very few imprinted genes have been identified in some livestock species (cattle, sheep, pig and rabbit) to be associated with the economically important traits such as milk yield, fat and meat deposition, fetal development, growth and carcass traits. Although the bovine genome is the best characterized livestock genome with high sequence coverage and with the highest percentage of annotated genes, however, less than two dozen imprinted genes have been experimentally validated (Imumorin et al., 2012). Imprinted genes are characterized by some genetic and epigenetic features. A remarkable feature of imprinted genes is that they are physically linked in clusters with other imprinted genes and do show unique patterns of sequence conservation (Hutter et al., 2010). The DNA sequence environment of imprinted genes is usually rich in CpG islands (CGI), repetitive elements and transcription factor binding sites (TFBS) (Paulsen et al., 2000; Neumann et al., 1995). These features are being used to analyze known and putative imprinted genes (Khatib et al., 2007; Luedi et al., 2007). Recently, a large cluster of imprinted genes has been mapped to the bovine chromosome 29 (Imumorin et al., 2012). Bovine chromosome (Bta) 29 is the equivalent of human chromosome

(Hg) 11 and mouse chromosome (Mm) 7 which contains the highest number of imprinted genes. In the current study, we selected a total of 14 genes from the imprinting cluster on Bta 29 and carried out comparative genomic analysis with its human and mouse orthologs to facilitate a better understanding of the imprinting sequence features which will aid in the experimental validation of their respective imprinting status.

2. Materials and Method

2.2 In silico Sequence Retrieval

The fourteen orthologous genes selected from the imprinting clusters on Bta 29, Hg 11 and Mm 7 were *H19*, *IGF2*, *INS*, *TH*, *ASCL2*, *TSPAN32*, *CD81*, *TSSC4*, *KCNQ1*, *CDKN1C*, *SLC22A18*, *PHLDA2*, *NAP1L4* and *OSBPL5*. A structured query of these genes in the Otago catalogue of imprinted genes (http://www.otago.ac.nz/IGC) was performed to identify the respective imprinting status across the three species (Table 1). Genomic, transcript and protein reference sequences (RefSeq) for each of the genes were retrieved from the GenBank (NCBI).

Table 1. Sizes (kb) and imprinting status of the Bta 29 genes and its orthologs in human and mouse

Gene Symbol	Name	Cattle		Human		Mouse	
		Size*	Status	Size*	Status	Size*	Status
H19	Imprinted maternally expressed transcript	2.37	ME	2.66	ME	2.62	ME
IGF2	Insulin-like growth factor 2	18.61	PE	20.49	PE	10.27	PE
INS	Insulin	1.16	Unknown	1.43	PE	1.05	PE
TH	Tyrosine hydroxylase	6.86	Unknown	7.88	Unknown	7.19	ME
ASCL2	Achaete-scute complex homolog 2	1.88	Unknown	2.46	ISE	2.44	ME
TSPAN32	Tetraspanin 32	48.61	Unknown	16.19	NI	14.44	ME
CD81	CD81 molecule	6.17	Unknown	20.1	NI	15.18	ME
TSSC4	Tumor suppressing subtransferable candidate 4	3.64	ISE	1.58	NI	1.72	ME
KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1	327.62	Unknown	404.12	ME	319.79	ME
CDKN1C	Cyclin-dependent kinase inhibitor 1c	2.01	Unknown	2.55	ME	2.7	ME
SLC22A18	Solute carrier family 22, member 18	19.07	Unknown	25.53	ME	25.58	ME
PHLDA2	Pleckstrin homolog-like domain, family A member 2	1.06	ME	1.15	ME	0.977	ME
NAP1L4	Nucleosome assembly protein 1-like 4	52.28	Unknown	47.95	Unknown	35.5	ISE
OSBPL5	Oxysterol binding protein-like 5	35.58	Unknown	78.24	ME	53.2	ME

Note. ME = maternally expressed; PE = paternally expressed; ISE = insufficient evidence; NI = not imprinted; Unknown = record unavailable; * = sizes in kilo base.

Source: Otago catalogue of imprinted genes.

2.1 Bioinformatics Analysis

Computational prediction of CGI was performed using four web-based programmes (CpGIS Searcher: http://cpgislands.com; CpGPROD: http://pbil.univ-lyon1.fr/software/cpgprod.html; CpGREPORT: http://emboss.sourceforge.net/apps/cvs/emboss/apps/cpgreport.html and CpGPLOT: http://www.ebi.ac.uk/tools/emboss/cpgplot.html). The genomic sequences were then submitted to all the CGI prediction programmes. The default prediction criteria of the programmes were maintained. The output generated from the CGI prediction of the genes was collated and analyzed across the prediction algorithms as described by Hackenberg et al. (2006, 2011) and Bock et al. (2006). RepeatMasker (http://www.repeatmasker.org) was used to screen for the genomic interspersed repeats. The transcription factor binding elements were determined using two programmes; rVista 2.0 and TFBIND which are freely available at http://rvista.dcode.org and http://tfbind.hgc.jp respectively.

MEME (http://meme.nbcr.net) was used to identify the conserved sequence motifs in the protein RefSeq. The functional and structural domain of the identified motifs was determined using the Pratt and ScanProsite programmes (http://www.expasy.org). The identified domains were further queried for specific residue patterns using the Motif scan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

2.2 Phylogenetic Analysis

The genomic RefSeq was used to determine the nucleotide variations; the coding sequence (CDS) was used to test for the type of selection and the protein sequence was used to build the phylogenetic trees and as well as the distance matrix. Using the pairwise sequence comparison (*i.e.* cattle/human, cattle/mouse, human/mouse), the rate of non-synonymous substitutions (d_N) and synonymous substitutions (d_S), d_N/d_S as well as neutrality index (NI) were determined from the CDS. These estimates were obtained using the DnaSP 5.0 programme (Librado & Rozas, 2009). Protein sequences were used to estimate the evolutionary distance and subsequently infer the phylogenetic relationships. The phylogenetic trees were constructed using the Neighbor-Joining method of MEGA5.2 software (Tamura et al., 2007) by selecting the pairwise distance model for amino acids substitutions. A 1000 bootstrap replication test was performed and the evolutionary distances (p-distance matrix) formed the nodal parameter used in defining the clades.

3. Results

3.1 DNA Sequence Features

The CGI predicted by each of the programmes for the putative imprinted genes on Bta 29 as well as its human and mouse orthologs were classified into promoter, intragenic and gene-terminal CGI (Figure 1).

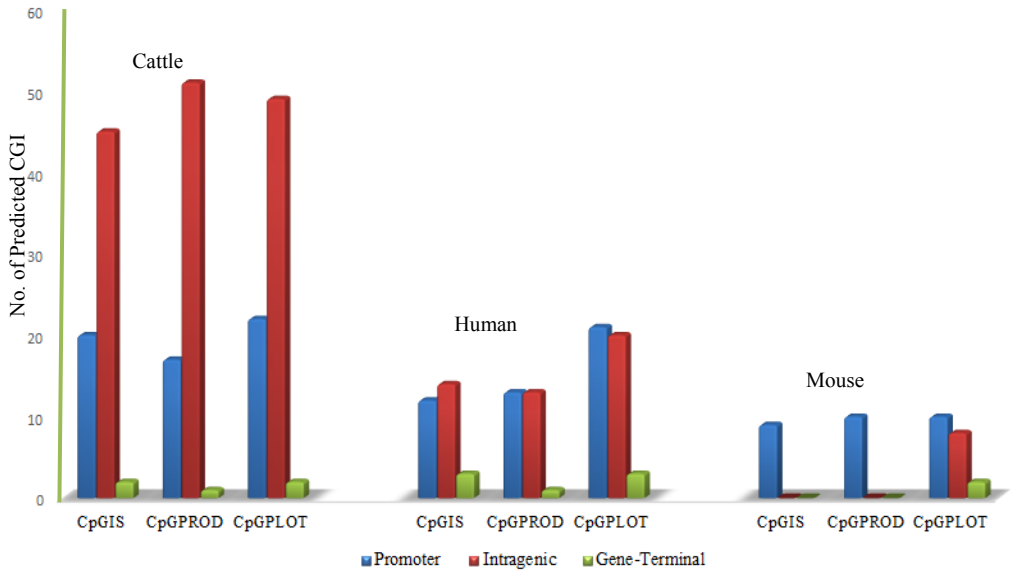

Figure 1. CGI categories across the finders for cattle, human and mouse

The CGI predicted by CpGPLOT and CpGREPORT were similar. Only six (*ASCL2*, *TSSC4*, *KCNQ1*, *CDKN1C*, *PHLDA2* and *NAP1L4*) out of the 14 genes in the imprinted gene clusters on Bta 29, Hg 11 and Mm 7 were identified to have promoter CGI across each of the prediction programmes. Also, across the various programmes, no CGI were predicted for two genes in cattle (*INS*, *CD81*), two in human (*INS*, *TSPAN32*) and five in mouse (*INS*, *TH*, *TSPAN32*, *SLC22A18* and *OSBPL5*). There were no CGI found in *INS*.

The repetitive elements were analyzed to compare the frequency and distribution of the short interspersed nuclear elements (SINE) particularly the *Alu* (*Arthrobacter luteus*) repeats. Table 2 shows there were repetitive sequences in all the 14 genes across the three species except for *OSBPL5* and *H19* in cattle and human respectively. There were no *Alu* repeats in the repetitive elements of all the putative imprinted genes on Bta 29.

Table 2. Percentage of repetitive sequences in Bta 29 imprinted gene cluster and its orthologs in human and mouse

Genes	Bta 29			Hg 11			Mm 7		
	BM %	SINE%	ALU%	BM %	SINE%	ALU%	BM %	SINE%	ALU%
H19	0.53	-	-	-	-	--	0.80	-	-
IGF2	4.89	0.77	-	6.72	-	-	2.59	-	-
INS	7.37	6.55	-	3.20	3.18	-	3.50	3.54	3.54
TH	2.98	-	-	2.40	0.88	-	4.03	0.66	-
ASCL2	7.79	2.92	-	8.22	2.72	-	9.90	6.70	-
TSPAN32	54.99	12.72	-	9.52	1.14	0.80	12.87	5.49	0.69
CD81	2.18	0.61	-	12.11	1.84	0.70	6.25	2.45	1.36
TSSC4	19.20	17.09	-	3.82	3.82	-	1.10	-	-
KCNQ1	40.83	10.41	-	41.20	5.10	3.80	30.00	5.72	1.08
CDKN1C	8.93	-	-	17.60	-	-	12.25	-	-
SLC22A18	8.43	4.59	-	19.70	4.50	3.20	20.03	9.19	3.65
PHLDA2	5.17	-	-	6.50	1.68	-	26.27	20.83	3.16
NAP1L4	26.98	13.57	-	28.50	19.30	18.20	25.90	19.38	9.78
OSBPL5	-	-	-	31.00	9.60	6.60	19.90	10.80	4.20

Note. BM = bases masked; ALU = *Arthrobacter luteus* expressed as a percentage of the SINE; SINE = short interspersed transposable elements; HG = human; MM = mouse.

The promoter CGI across the six genes were enriched with the CpG-containing E2F, ZF, EGR, KROX, SP1, AP2 and YY1 transcription factor binding sites (consensus sequence) but there were no TATA boxes. Evolutionary conserved domains were identified in all the six genes but site-specific motifs were only found in *ASCL2*, *TSSC4*, *PHLDA2* and *NAP1L4* (Table 3). These motifs were found within the conserved protein domain of each of the respective genes. The motifs were active sites for protein kinases involved in phosphorylation which is an important epigenetic mechanism for post-translation modification of gene expression

Table 3. Evolutionary conserved domains and motifs

Genes	Species	Genomic domain	Motif sites*	Protein kinase
ASCL2	Cattle/human	BHLH	83-85	PKC
		50-102	84-87	cAMP
	Mouse	BHLH	87-90	CK2
		118-170	91-93	PKC
			94-97	CK2
			152-155	cAMP
			155-158	CK2
			159-161	PKC
			162-165	CK2
TSSC4	Cattle	undefined	86-88	PKC
			91-94	CK2
	Human	undefined	90-92	PKC
	Mouse	undefined	87-89	PKC
			92-96	CK2
KCNQ1	Cattle/human/mouse	KCNQ1 Channel	-	-
CDKN1C	Cattle	CDI 20-72	-	-
	Human	CDI 31-83		
	Mouse	CDI 33-85		
PHLDA2	Cattle	PH 7-99	64-66	PKC
	Human	PH 7-99	75-77	PKC
	Mouse	PH 18-109	64-66	PKC
			74-76	PKC
NAP1L4	Cattle/human/mouse	NAP	315-318	CK2
		64-341		

Note. BHLH = myogenic basic helix-loop-helix; IT = ion transport; CDI = cyclin-dependent kinases inhibitor; PH = pleckstrin homology; NAP = nucleosome assembly protein; PKC = protein kinase C; cAMP = cyclic adenosine monophosphate; CK2 = casein kinase C; * = amino-acids residue position.

3.2 Molecular Evolution

The analyses of the genetic diversity was with respect to three pairwise datasets. Transition bias (Ts>Tv) which is a general property of DNA sequence evolution was observed in *ASCL2*, *TSSC4* and *CDKN1C* (Table 4). In *CDKN1C*, all the pair-wise comparisons showed transition bias whereas in *ASCL2* and *TSSC4*, the phenomenon only occurred in two of the pairwise comparisons. The transition/transversion rate ratio as estimated by the distance-based and maximum likelihood methods was less than 2 (Table 5). The pairwise comparison between cattle and human had the lowest nucleotide diversity (P*i*) across the six genes except in *NAP1L4* (Table 4).

Table 4. Phylogenetic analysis of the six putative imprinted bovine genes

Genes	Parameter	All 3 spp.	Cattle/human	Cattle/mouse	Human/mouse
ASCL2	T*s*	257	216	283	371
	T*v*	207	197	263	372
	P*i*	0.21	0.23	0.32	0.33
TSSC4	T*s*	288	245	312	306
	T*v*	288	200	405	258
	P*i*	0.36	0.30	0.43	0.39
KCNQ1	T*s*	72825	70109	72958	75409
	T*v*	85578	76534	93966	86233
	P*i*	0.53	0.48	0.59	0.52
CDKN1C	T*s*	290	196	269	307
	T*v*	277	154	201	265
	P*i*	0.27	0.18	0.25	0.23
PHLDA2	T*s*	172	166	174	175
	T*v*	208	185	217	222
	P*i*	0.38	0.35	0.41	0.42
NAP1L4	T*s*	9924	11803	9583	8386
	T*v*	11911	13220	13054	9460
	P*i*	0.57	0.53	0.65	0.52

Note. T*s* = transitions; T*v* = transversions; P*i* = nucleotide diversity.

Table 5. Estimates of the mutational transition/transversion rate ratio

Genes	Transition/transversion rate ratio		
	Distance-based model (κ=P/Q)	Maximum likelihood models (κ=α/β)	
		HKY	Kimura 2 parameter
ASCL2	1.25	1.40	1.34
TSSC4	1.00	1.47	1.39
KCNQ1	0.85	1.29	1.29
CDKN1C	1.05	1.58	1.49
PHLDA2	0.83	1.05	1.01
NAP1L4	0.83	1.37	1.39

Note. P = transitional sites; Q = transversional sites; α = instantaneous transition rate; β = instantaneous transversion rate; HKY = Hasegawa-Kishino-Yano.

The d_N/d_S rate ratio was found to be less than 1 ($d_N/d_S < 1$) for all the pairwise comparisons in *ASCL2*, *TSSC4*, *CDKN1C* and *NAP1L4*. *KCNQ1* had a $d_N/d_S > 1$ in all the pairwise comparisons while *PHLDA2* had a $d_N/d_S < 1$ in all the pairwise comparisons except cattle/human (Table 6). The phylogenetic trees (Figure 2) showed that the out-group (Zebrafish/Red Jungle Fowl) was classified differently from the mammals which is consistent with the traditional classification. The results showed that cattle was more closely related to human than mouse in all the six genes except for *NAP1L4*.

Table 6. Synonymous (d_S) and Nonsynonymous (d_N) nucleotide substitution rate

Genes	Parameter	Substitutions	d_S	d_N	d_N/d_S	NI
ASCL2	Cattle/human	93	0.451	0.093	0.206	0.498
	Cattle/mouse	145	0.754	0.183	0.243	0.840
	Human/mouse	127	0.492	0.185	0.376	1.094
TSSC4	Cattle/human	222	0.588	0.207	0.352	0.790
	Cattle/mouse	281	1.029	0.253	0.246	1.112
	Human/mouse	262	0.964	0.226	0.234	1.157
KCNQ1	Cattle/human	189	0.106	0.140	1.320	1.263
	Cattle/mouse	236	0.138	0.197	1.428	1.024
	Human/mouse	222	0.116	0.190	1.638	1.070
CDKN1C	Cattle/human	142	0.460	0.172	0.374	0.736
	Cattle/mouse	180	0.481	0.270	0.561	1.015
	Human/mouse	172	0.469	0.250	0.533	0.821
PHLDA2	Cattle/human	292	0.405	0.471	1.163	0.758
	Cattle/mouse	349	0.635	0.565	0.890	1.770
	Human/mouse	363	0.635	0.614	0.967	1.680
NAP1L4	Cattle/human	145	0.847	0.028	0.033	0.204
	Cattle/mouse	142	0.805	0.028	0.035	0.211
	Human/mouse	159	1.116	0.023	0.021	0.104

Note. NI = neutrality index.

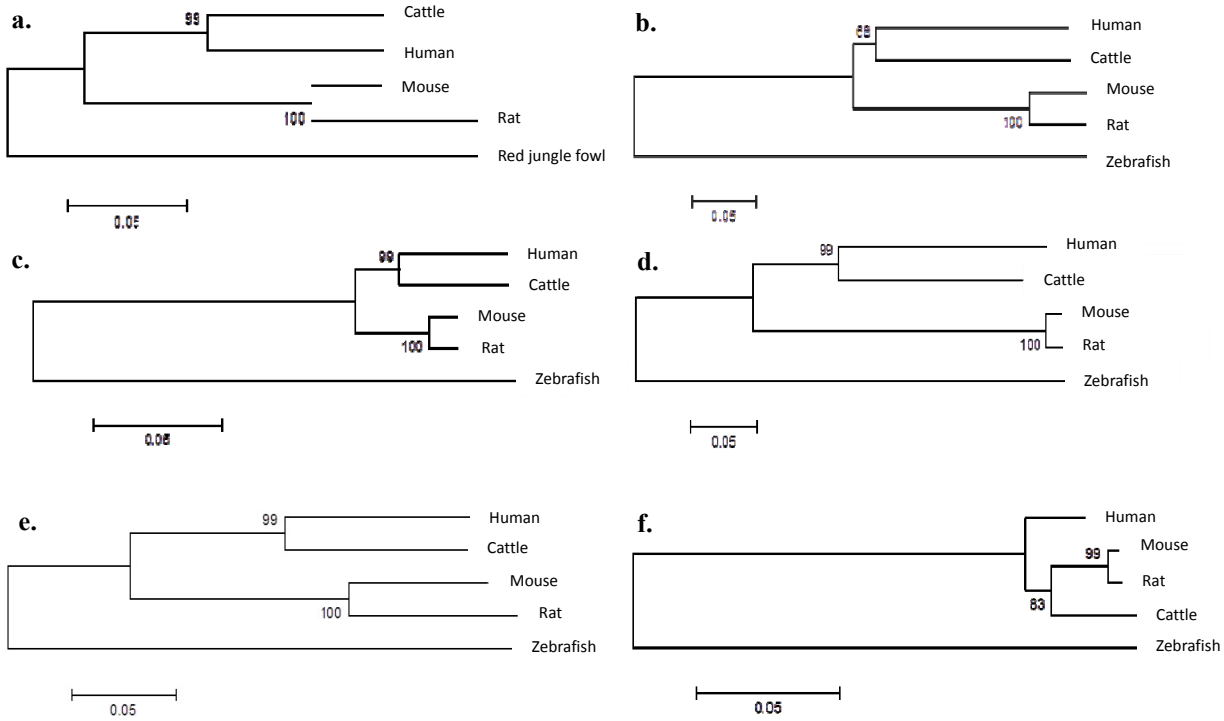

Figure 2. The phylogenetic relationships between cattle and other mammals using Zebrafish or Red jungle fowl as the out-group

Note. a = ASCL2, b = TSSC4, c = KCNQ1, d = CDKN1C, e = PHLDA2, f = NAP1L4.

4. Discussion

The analysis of the repetitive elements showed that the mammalian-wide interspersed repeats (MIRs) accounted for the SINE transposons as there were no *Alus* in all the 14 putative imprinted genes in cattle. This is expected since *Alus* are primate specific SINEs (Liu et al., 2009). Imprinted loci have been reported to contain fewer SINE transposons-derived sequences than non-imprinted loci and that there is a direct relationship between SINEs and imprinting (Greally, 2002). *NAPIL4* had an unusual high percentage of SINEs in human. This according to Greally (2002), is characteristic of the sequence transition region *TSSC3*/*NAPIL4* which is flanked by regions of increased SINE content. It was observed that the mouse orthologs were CpG poor which supports earlier reports in which about 20% of mouse orthologs of human genes do not always have CGI as a result of the evolutionary pressure towards conservation (Antequera, 2003; Illingworth et al., 2010).

The predicted CGI were assigned into promoter, intragenic and gene-terminal CGIs as described by Bock et al. (2006). Intragenic CGI also described as 'orphan CGIs' by Illingworth et al. (2010), is said to play a role in transcriptional initiation and dynamic expression during development. As such the abundance of these orphan CGI in cattle and human suggests a regulatory function associated with the various isoforms of the putative imprinted genes. The identification of promoter CGI in *ASCL2*, *TSSC4*, *KCNQ1*, *CDKN1C*, *PHLDA2* and *NAP1L4* is significant because promoter CGI are the major sequence characteristics of imprinted genes (Paulsen et al., 2008). Also, genes with promoter CGI often function as housekeeping genes (Weber et al., 2007). Three out of these genes (*KCNQ1*, *CDKN1C* & *PHLDA2*) had been experimentally validated to be imprinted in human while the other three are imprinted in mouse (Morison et al., 2005). Recently, *PHLDA2* was reported to be imprinted in cattle (Sikora et al., 2012). This therefore suggests that *ASCL2*, *TSSC4*, *KCNQ1*, *CDKN1C*, and *NAP1L4* may also be imprinted in cattle and their promoter CGI functionally involved in differential gene expression.

Our result supports earlier studies in which several TFs have been reported to contain CpGs in their recognition sequence and that promoter CGI lack TATA boxes (Deaton & Bird, 2011). According to Landolin et al. (2010), the enrichment of the promoter CGI with CpG-containing transcription factor binding sites is characteristic of imprinted genes. In this study, we identified several regions of sequence conservation in which core TF binding

sites (Sox2, Nanog, Oct4) were found. Although most of the identified TFs were present within UTR and intronic regions, however, these may be potential sites for differential methylation (Hansen et al., 2012). A cross-matching of the identified conserved intergenic regions in five (*INS*, *TH*, *ASCL2*, *TSSC4* and *PHLDA2*) of the putative imprinted genes corresponds to the imprinted gene clusters on Bta 29. This suggests that the core TFs within these intergenic regions may provide additional regulatory signals for the respective imprinting control centers (IC) of the gene clusters (Paulsen et al., 2008). The observed transition bias in *ASCL2***,** *TSSC4* and *CDKN1C* indicates that during the speciation of these genes, transition base substitutions were favoured over transversions inorder to ensure the conservation of the chemical nature of the proteins (Wakeley, 1996). According to Zhao et al. (2006), base substitution is the main cause of gene variation, diversity and evolution of species. For all the six genes, the estimates of the mutational transition/transversion rate ratio were less than two (< 2) which according to Wang et al. (2012), suggests that mutations within these genes in the mammalian homologs (cattle, human & mouse) have reached a saturation status.

In this study, all the six genes except *KCNQ1* had a d_N/d_S that is significantly less than one ($d_N/d_S < 1$; Z-test, $p < 0.05$) and with significant deviation from the neutral theory (NI < 1; Fisher's exact test, $p < 0.05$). This demonstrates that the evolution of *ASCL2*, *TSSC4*, *CDKN1C*, *PHLDA2* and *NAP1L4* have been driven by natural (negative) selection and not random drift. The constraining of d_N is a way by which natural selection prevents potential changes to the underlying amino-acids, thereby stabilizing the expression of the respective gene products (Wolf et al., 2009). According to Schaffner and Sabeti (2008), diet, climate and disease are the most significant forces driving the conservation of amino-acid residues in mammalian populations.

The phylogenetic trees for *ASCL2*, *TSSC4*, *CDKN1C*, *KCNQ1* and *PHLDA2* were consistent with earlier studies wherein bovine proteins were reported to share more homology with humans than mouse (Tellam et al., 2009). According to Tellam et al. (2009), alterations in the organization of specific gene families in the bovine lineage could have informed peculiar genome similarities and differences across other mammalian species. It thus suggests that the evolution of all these six putative imprinted genes except *NAP1L4*, may have undergone cattle-specific changes that are indicative of the evolutionary adaptations to the immediate environment, disease challenges (Elsik et al., 2009), reproductive functions (Rodriguez-osorio et al., 2009), growth and development (Ulzun et al., 2009). The conserved site-specific motifs in *ASCL2*, *TSSC4*, *PHLDA2* and *NAP1L4* were protein kinases involved in phosphorylation. The role of protein kinases in phospho-regulation has been compared to the transcription regulatory activity of TFs. According to Mair (2009), just as TFs regulate genes via recognizing specific DNA sequences, protein kinases only phosphorylate proteins that contain particular amino acid motifs. The specific role of each of these protein kinases (cAMP, PKC, CK2) in glycogen regulation, muscle development and cellular regulatory mechanism within the respective domains of the queried genes could be an important evolutionary source of phenotype variability (Beltrao et al., 2009). The absence of any site-specific motif in the conserved domains of the bovine putative imprinted genes; *KCNQ1* and *CDKN1C* as well as in its human and mouse orthologs suggests that the imprinting mechanism of these genes lie solely within its DNA sequences. That is, phospho-regulation may not be involved in the epigenetic regulation of the two genes.

5. Conclusion

The *in silico* characterization of the imprinted genes can then be used to predict molecular function or gene expression. The six bovine putative imprinted genes identified to have promoter CGI can be further assayed to determine their DNA methylation status. This study confirms that at the proteomic level the bovine genome shares more homology with humans than mouse which is consistent with the National Human Genetic Research Institute's assessment of cattle as an excellent model species for biomedical research. Our study reports the *in silico* phospho-regulatory mechanism of imprinting in *ASCL2*, *TSSC4*, *PHLDA2* and *NAP1L4*. This post-translation modification will require further experimental validation especially in genes that have defied the DNA methylation hypothesis for genomic imprinting.

Acknowledgements

The study is a portion of the Ph.D. thesis of the first author. The work was jointly supported by the Department of Animal Sciences, Obafemi Awolowo University Ile-Ife, Nigeria and College of Agriculture and Life Sciences, Cornell University, Ithaca, NY.

References

Antequera, F. (2003). Structure, function and evolution of CpG Island promoters. *Cell and Molecular Life Science, 60*(8), 1647-1658. http://dx.doi.org/10.1007/s00018-003-3088-6

Beltrao, P., Trinidad, J. C., Fiedler, D., Roguev, A., Lim, W. A., Shokat, K. M., ... Krogan, N. J. (2009). Evolution

of Phosphoregulation: Comparison of Phosphorylation Patterns across Yeast Species. *PLoS Biol, 7*(6), e1000134. http://dx.doi.org/10.1371/journal.pbio.1000134

Blankenberg, D., Taylor, J., Schenck, I., He, J., Zhang, Y., Ghent, M., ... Nekrutenko, A. A. (2007). Framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. *Genome Research, 17*(6), 960-4. http://dx.doi.org/10.1101/gr.5578007

Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T., & Walter, J. (2006). CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats and predicted DNA structure. *Public Library of Science Genetics, 2*(3), e26. http://dx.doi.org/10.1371/journal.pgen.0020026

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development, 25*, 1010-1022. http://dx.doi.org/10.1101/gad.2037511

Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Muzny, D. M., Weinstock, G. M., ... The Bovine Genome Sequencing and Analysis Consortium. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science, 324*(5926), 522-528. http://dx.doi.org/10.1126/science.1169588

Garfield, A. S., Cowley, M., Smith, F. M., Moorwood, K., Stewart-Cox, J. E., Gilroy, K., ... Ward, A. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature, 469*, 534-538. http://dx.doi.org/10.1038/nature09651

Greally, J. M. (2002). Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proceedings of the National Academy of Science, 99*(1), 27-32. http://dx.doi.org/10.1073/pnas.012539199

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., & Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science, 329*(5992), 643-8. http://dx.doi.org/10.1126/science.1190831

Hackenberg, M., Carpena, P., Bernaola-Galván, P., Barturen, G., Alganza, A. M., & Oliver, J. L. (2011). WordCluster: detecting clusters of DNA words and genomic elements. *Algorithm for Molecular Biology, 6*, 2. http://dx.doi.org/10.1186/1748-7188-6-2

Hackenberg, M., Previti, C., Luque-Escamilla, P. L., Carpena, P., Martinez-Aroza, J., & Oliver, J. L. (2006). CpGcluster: A distance-based algorithm for CpG-island searching. *BioMed Central Bioinformatics, 7*, 446.

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology, 13*, R83. http://dx.doi.org/10.1186/gb-2012-13-10-r83

Hutter, B., Bieg, M., Helms, V., & Paulsen, M. (2010a). Imprinted genes show unique patterns of sequence conservation. *BioMed Central Genomics, 11*, 649. http://dx.doi.org/10.1186/1471-2164-11-649

Illingworth, R. S., & Bird, A. P. (2009). CpG islands- 'a rough guide'. *Federation of European Biochemical Societies Letters, 583*, 1713-1720. http://dx.doi.org/10.1016/j.febslet.2009.04.012

Imumorin, I. G., Peters, S. O., & De Donato, M. (2012). Genomic imprinting and imprinted gene clusters in the bovine genome. In H. Khatib (Ed.), *Livestock Epigenetics* (pp. 89-112). Wiley-Blackwell. http://dx.doi.org/10.1002/9781119949930.ch6

Kedia-Mokashi, N. A., Mugasimangalam, R., Aiyaz, M., Mukherjee, S., & Balasinor, N. H. (2011). Aberrant expression of imprinted genes in post-implantation rat embryos. *Life Sciences, 88*(13), 634-643. http://dx.doi.org/10.1016/j.lfs.2011.02.005

Khatib, H., Zaitoun, I., & Kim, E. S. (2007). Comparative analysis of sequence characteristics of imprinted genes in human, mouse and cattle. *Mammalian Genome, 18*(6-7), 538-47. http://dx.doi.org/10.1007/s00335-007-9039-z

Landolin, J. M., Johnson, D. S., Trinklein, N. D., Aldred, S. F., Medina, C., Shulha, H., ... Myers, R. M. (2010). Sequence features that drive human promoter function and tissue specificity. *Genome Research, 20*, 890-898. http://dx.doi.org/10.1101/gr.100370.109

Lim, D. H. K., & Maher, E. R. (2010). DNA methylation: a form of epigenetic control of gene expression. *The Obstetrician & Gynaecologist, 12*, 37-42. http://dx.doi.org/10.1576/toag.12.1.037.27556

Liu, Y., Qin, X., Song, X., Jiang, H., Shen, Y., Durbin, K. J., ... Gibbs, R. A. (2009). Bos taurus genome assembly. *BioMed Central Genomics, 10*, 180. http://dx.doi.org/10.1186/1471-2164-10-180

Luedi, P. P., Dietrich, F. S., Weidman, J. R., Bosko, J. M., Jirtle, R. L., & Hartemink, A. J. (2007). Computational and experimental identification of novel human imprinted genes. *Genome Research*. http://dx.doi.org/10.1101/gr.6584707

Mair, W. (2009). Change for the Better: Phosphoregulation of Proteins Drives Evolution. *Public Library of Science Biology, 7*(6), e1000127. http://dx.doi.org/10.1371/journal.pbio.1000127

Morison, I. M., Ramsay, J. P., & Spencer, H. G. (2005). A census of mammalian imprinting. *Trends in Genetics, 21*, 457-465. http://dx.doi.org/10.1016/j.tig.2005.06.008

Neumann, B., Kubicka, P., & Barlow, D. P. (1995). Characteristics of imprinted genes. *Nature Genetics, 9*, 12-13. http://dx.doi.org/10.1038/ng0195-12

Nikaido, I., Saito, C., Mizuno, Y., Meguro, M., Bono, H., Kadomura, M., ... GSL Members. (2003). Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Research, 13*(6B), 1402-9. http://dx.doi.org/10.1101/gr.1055303

Paulsen, M., El-Maarri, O., Engemann, S., Strodicke, M., Franck, O., Davies, K., ... Walter, J. (2000). Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Human Molecular Genetics, 9*, 1829-1841. http://dx.doi.org/10.1093/hmg/9.12.1829

Paulsen, M., Tierling, S., & Walter, J. (2008). DNA methylation and the mammalian genome. In J. Tost (Ed.), *Epigenetics* (pp. 1-21). Caister Academic Press.

Ponger, L., & Mouchiroud, D. (2002). CpGProD: Identifyng CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics Application Note, 18*(4), 631-633. http://dx.doi.org/10.1093/bioinformatics/18.4.631

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics, 16*(6), 276-277. http://dx.doi.org/10.1016/S0168-9525(00)02024-2

Rodriguez-Osorio, N., Wang, Z., Kasinathan, P., Page, G. P., Robl, J. M., & Memili, E. (2009). Transcriptional reprogramming of gene expression in bovine somatic cell chromatin transfer embryos. *BioMed Central Genomics, 10*, 190. http://dx.doi.org/10.1186/1471-2164-10-190

Schaffner, S., & Sabeti, P. (2008). Evolutionary adaptation in the human lineage. *Nature Education, 1*(1), 14.

Schulz, R., Proudhon, C., Bestor, T. H., Woodfine, K., & Lin, C. S. (2010). The Parental Non-Equivalence of Imprinting Control Regions during Mammalian Development and Evolution. *Public Library of Science Genetic, 6*(11), e1001214. http://dx.doi.org/10.1371/journal.pgen.1001214

Sikora, K. M., Magee, D. A., Berkowicz, E. W., Lonergan, P., Evans, A. C. O., Carter, F. A., ... Spillane, C. (2012). PHLDA2 is an imprinted gene in cattle. *Animal Genetics, 43*(5), 587-590. http://dx.doi.org/10.1111/j.1365-2052.2011.02292.x

Takai, D., & Jones, P. A. (2003). The CpG Island Searcher: A new WWW resource. *In silico Biology, 3*, 235-240.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software. *Molecular Biology Evolution, 24*, 1596-1599. http://dx.doi.org/10.1093/molbev/msm092

Tellam, R. L., Lemay, D. G., Van Tassell, C. P., Lewin, H. A., Worley, K. C., & Elsik, C. G. (2009). Unlocking the bovine genome. *BioMed Central Genomics, 10*, 193. http://dx.doi.org/10.1186/1471-2164-10-193

Ulzun, N., Rodriguez-Osorio, A., Kaya, H., Wang, J. J., Parrish, V. A., Ilyin, E., & Memili, E. (2009). Functional Genomics of HMGN3a and SMARCAL1 in Early Mammalian Embryogenesis. *BioMed Central Genomics, 10*, 183.

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution, 11*, 158-162. http://dx.doi.org/10.1016/0169-5347(96)10009-4

Wang, S., Pan, Z., Zhang, Q., Xie, Z., Liu, H., & Li, Q. (2012). Differential mRNA expression and promoter methylation status of SYCP3 gene in testes of Yaks and cattle-Yaks. *Reproduction in Domestic Animal, 47*, 455-462. http://dx.doi.org/10.1111/j.1439-0531.2011.01902.x

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., & Paabo, S. (2007). Distribution, siliencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics, 39*, 457-466. http://dx.doi.org/10.1038/ng1990

Wolf, J. B. W., Kunstner, J., Nam, K., Jakobsson, M., & Ellegren, H. (2009). Nonlinear Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects Inference of Selection. *Genome Biology & Evolution, 1*, 308-319. http://dx.doi.org/10.1093/gbe/evp030

Zhao, H., Qizhai, L., Li, J., Zeng, C., Hu, S., & Yu, J. (2006). The study of neighboring nucleotide composition and transition/transversion bias. *Science in China*: *Series C Life Sciences, 49*(1), 1-8.