Predicting CpG Islands and Their Relationship with Genomic Feature in Cattle by Hidden Markov Model Algorithm

Research Article

A. Barazandeh[1,2], M.R. Mohammadabadi[1*], M. Ghaderi-Zefrehei[3] and H. Nezamabadipour[4]

[1] Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran
[2] Department of Animal Science, University of Jiroft, Jiroft, Iran
[3] Department of Animal Science, University of Yasouj, Yasouj, Iran
[4] Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

*Correspondence E-mail: mrm@uk.ac.ir

ABSTRACT

Cattle supply an important source of nutrition for humans in the world. CpG islands (CGIs) are very important and useful, as they carry functionally relevant epigenetic loci for whole genome studies. As a matter of fact, there have been no formal analyses of CGIs at the DNA sequence level in cattle genomes and therefore this study was carried out to fill the gap. We used hidden markov model algorithm to detect CGIs. The total number of predicted CGIs for cattle was 90668. The number of detected CGIs and CGI densities downwardly varied across chromosomes. Chromosome 25 had the largest number of CGIs (4556) and the highest CGI density (106.20 CGIs/Mb). A significant positive correlation observed among CGI densities with guanine-cytosine (GC) content, Obs_{CpG}/Exp_{CpG}, recombination rate and gene density. When the size of chromosomes increased, the CGI densities decreased and a trend of higher CGI densities in the telomeric regions observed. This feature may be the reason of a positive correlation between CGI density and recombination rate. To detect information on CGI density differences between cattle and other vertebrate genomes, CGI density was also scanned in eleven vertebrate genomes. The CGI densities varied greatly among genomes. These discoveries may contribute to a better understanding of epigenomic role of CGIs and their molecular evolution in the cattle.

KEY WORDS cattle, CpG islands, epigenomic, genome, HMM.

INTRODUCTION

Cattle provide a significant source of nutrition and livelihood to humans in the world. Cattle belong to Ruminantia, which occupy diverse terrestrial environments. They are renowned for their ability to efficiently convert low quality forage into energy, fat, muscle and milk. These biological processes have been exploited by the human species since domestication, which began in the Near East some 8000-10000 years ago (Willham 1986; Elsik *et al.* 2009). Since then, over 800 cattle breeds have been established representing an important world heritage and a scientific re-

source for understanding the genetics of complex traits. DNA methylation is a type of chemical modification of DNA that can be inherited without changing the DNA sequence (Amiri Roudbar *et al.* 2015). It involves the addition of a methyl group to DNA and typically occurs at CpG dinucleotide (Vercelli, 2016) and it is an important part of gene regulation (Irizarry *et al.* 2009; Wu *et al.* 2010). The DNA of most vertebrates, especially mammals, is depleted in CpG dinucleotides. The remaining CpGs are clustered in regions referred to as CpG islands (CGIs). Interest in CGI has grown up because they are enriched in promoters of genes (Hackenberg *et al.* 2010) and by the presence of al-

tered DNA methylation in CGIs, they play important roles in the regulation of gene expression and gene silencing during processes such as X-chromosome inactivation, imprinting and silencing of intragenomic parasites (Takai and Jones, 2002; Su *et al.* 2010). Therefore, full comprehension of such CGI could help in discovering the epigenetic causes of cancer (Han and Zhao, 2009; Wu *et al.* 2010; Saito *et al.* 2016). Because of their crucial roles, multiple algorithms (either species specific or overall ones) have been developed for identifying CGIs in the genomes.

The first algorithm was proposed by Gardiner-Garden and Frommer in 1987, in which CGI is featured as a region of at least 200 bp, with GC content greater than 50% and the observed-to-expected CpG ratio (Obs_{CpG}/Exp_{CpG}) greater than 0.6 (Gardiner-Garden and Frommer, 1987; Glass *et al.* 2007; Tan *et al.* 2011; Yang *et al.* 2016). Due to these CGI inherent features, many algorithms and applications were designed to work according to these three features and also some new algorithms which circumvent preceding algorithm's cavities were also introduced (Rice *et al.* 2000; Takai and Jones, 2002; Wang and Leung, 2004). These algorithms depend on cutoffs and leaves out important CpG clusters associated with epigenetic marks, relevant to development and disease and since they were mainly developed for humans genome studies, they were not applicable at all to non-vertebrate genomes (Irizarry *et al.* 2009) . To solve these problems Wu *et al.* (2010) proposed an alternative Hidden Markov model-based approach that permitted an extensible approach to detecting CGI. The main advantage of this approach over others is that it summarizes the evidence for CGI status as probability scores. This provides flexibility in the definition of a CGI and facilitates the creation of CGI lists for other species. The utility of this approach is demonstrated by generating the first CGI lists for invertebrates and can create CGI lists that substantially increases overlap with discovered epigenetic marks (Wu *et al.* 2010). Hidden Markov models (HMMs) have been shown to be very effective in representing biological sequences and they have been successfully used for modeling speech signals (Rabiner, 1989). Thus, HMMs have become increasingly popular in computational molecular biology and many state of the arts studies. From 2009 onwards, the results of whole genome sequencing projects in domestic cattle have been released (Elsik *et al.* 2009). There have been some reports in terms of comparisons of CGIs and their correlation with genomic features in some mammalian genomes (mostly human and mouse) and non-mammalian genomes (fish) in the literatures (Glass *et al.* 2007; Han *et al.* 2008; Han and Zhao, 2008; Irizarry *et al.* 2009; Wu *et al.* 2010), but to our knowledge, there hasn't been a separate research of CGIs and their correlation with genomic features in cattle. Therefore, the main objective of this study was to investigate the presence of CGIs at the DNA sequence level in cattle genomes and outline a comparison to other vertebrate's genomes.

MATERIALS AND METHODS

Genome sequences and other genome information
Assembled genome sequences of cattle were downloaded from the National Center for Biotechnology Information (NCBI). The statistic and other genome information at the chromosome level are given in Table 1. CGIs of sheep, goat, dromedary camel, bactrian camel and alpaca were predicted by HMM algorithm. CGIs analysis for other vertebrate genomes (human, mouse, dog, horse, chicken and zebrafish) were downloaded from website (www.rafa-lab.Jhsph.edu/CGI/) and predicted according to HMM algorithm by Wu *et al.* (2010). Data of recombination rate in cattle (window size, 1 Mb) were obtained from (Weng *et al.* 2014).

Algorithm for detection of CGIs
CGIs were identified based on HMM algorithm (Irizarry *et al.* 2009; Wu *et al.* 2010). The foundation of this algorithm is the stochastic modeling of bases in the genome. This algorithm assumes that each genome is divided into 2 states (CGI and baseline). In this algorithm the genome is divided into non-overlapping segments of length L (in bp). The length L= 16 is used for the segments, because at this length the association of identified CGI with epigenetic markers is higher, then use the number of C, G and CpG in segment of length L as parameters for the model and the hidden state Y(s) for segment s with states Y(s)= 1 as CGI and Y(s)= 0 as baseline. Assume that Y(s) is a stationary first-order Markov chain. The choice of the state is based on two HMM. One is for GC content to be high or low with assumption of the binomial distribution approximated with the normal density. The second one is for CpG number with the assumption of Poisson distribution as follow:

$$a_i \times L \times p(s)^2 / 4$$

Where:
a_i: O/E for the CGI (i=1) and baseline (i=0).
p(s): GC content for segment that is calculated in the first step.

As a final step, the algorithm could obtain posterior probabilities of being in each state and creates lists of CGI using different specificity cutoffs values. A cutoff value of 0.9 is chosen based on the association of CGI with epigenetic markers. For the mathematical details of this algorithm see Wu *et al.* (2010).

Table 1 Statistic and distribution of CpG islands and genome in chromosomes of cattle

Items	Whole genome	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9
Size (Mb)	2660.91	158.34	137.06	121.43	120.83	121.19	119.46	112.64	113.39	105.71
GC content in whole (%)	41.8	40.1	40.8	41.9	40.6	41.7	40.0	42.0	41.2	40.0
Number of gene	26,410	1,430	1,441	1,973	1,241	1,905	1,116	1,957	1,258	944
Number of CGIs	90668	3784	4082	3890	4042	4244	3831	3324	2233	3154
CGI density (/Mb)	34.47	23.90	29.78	32.03	33.45	35.02	32.07	29.51	19.69	29.84
GC content in CGIs (%)	63.30	62.66	63.18	63.84	63.29	63.58	63.63	64.16	62.97	62.39

Items	Chr10	Chr11	Chr12	Chr13	Chr14	Chr15	Chr16	Chr17	Chr18	Chr19
Size (Mb)	104.31	107.31	91.16	84.24	84.65	85.3	81.72	75.16	66.00	64.06
GC content in whole (%)	41.5	42.9	40.4	43.7	41.4	41.8	42.6	42.3	45.4	46.0
Number of gene	1,618	1,440	770	1,274	888	1,560	1,076	978	1,792	1,694
Number of CGIs	4248	4367	2746	2715	2019	2047	1933	3173	4371	4396
CGI density (/Mb)	40.72	40.70	30.12	32.23	23.85	24.00	23.65	42.22	66.23	68.62
GC content in CGIs (%)	62.82	64.14	61.97	62.95	63.35	63.31	63.71	63.69	63.58	63.55

Items	Chr20	Chr21	Chr22	Chr23	Chr24	Chr25	Chr26	Chr27	Chr28	Chr29	Chr X
Size (Mb)	72.04	71.6	61.44	52.53	62.71	42.9	51.68	45.41	46.31	51.51	148.82
GC content in whole (%)	41.0	43.1	43.4	43.4	41.9	47.1	42.8	41.8	42.2	44.2	40.6
Number of gene	601	967	850	1,231	540	1,043	638	484	510	968	1,791
Number of CGIs	2082	2383	2802	3236	2282	4556	2217	1698	678	2901	1234
CGI density (/Mb)	28.90	33.28	45.61	61.60	36.39	106.20	42.90	37.39	14.64	56.32	8.29
GC content in CGIs (%)	62.55	63.33	63.5	62.21	61.79	64.34	63.33	61.76	63.62	64.22	61.12

CGI: CpG island and GC: guanine-cytosine.

This algorithm is implemented as an R adds-in package for R software which is called make CGI.

CGI mapping to different genomic regions

The method of Han *et al.* (2008) is used to identify CGIs in different genomic regions (genes, intergenic regions, intragenic regions and transcriptional start sites (TSS) regions). Concisely, locations of CGIs compared with the coordinates of different genomic regions, is based on the cow gene annotation information from the NCBI and UCSC databases by BEDTools (Quinlan and Hall, 2010).

CGIs overlapped with any genes were classified as gene-associated CGIs; CGIs whose whole sequences were in intergenic regions were classified as intergenic CGIs; CGIs whose sequences were in gene regions were classified as intragenic CGIs and CGIs overlapped with TSSs were classified as TSS CGIs.

RESULTS AND DISCUSSION

CGIs and CGI density in cattle genome

CGIs detected by HMM algorithm for all chromosomes in cattle genome is summarized in Table 1. Because the size varied across chromosomes, the CGI density was measured by the average number of CGIs per Mb. The number of predicted CGIs and correspondingly CGI densities varied downwardly among chromosomes. Chromosome 25 had the largest number of CGIs (4556) and the highest CGI density (106.20 CGIs/Mb). The CGI density in the smallest chromosome (Chr 25) was about five times greater than the largest chromosome (Chr 1), that would give out a high density of CGIs on micro chromosomes. Moreover, previous analyses of CGIs in the chicken genome revealed a high concentration of CGIs on micro chromosomes (Hillier *et al.* 2004; Rao *et al.* 2013).

These results suggest that some other genomic factors might have also played important roles in the course of CGI evolution (Han and Zhao, 2008). Figure 1 shows box plots of the lengths of detected CGIs for all chromosomes. The average length and variance of CGI across chromosomes was low. These results were similar to other studies (Hackenberg *et al.* 2006; Chuang *et al.* 2011; Chuang *et al.* 2012).

The total number of predicted CGIs for cattle was 90668. Han *et al.* (2008) analyzed CGIs in 9 mammalian genomes. They predicted the number of CGIs in cattle to be 58327. Interestingly, the CGIs number reported there are quite lower than those reported in the current study, which is probably due to the different approaches used for predicting CGIs.

Correlation between CGI density and other genomic features of cattle genome

A significant negative correlation (r=-0.49, p=0.006) was observed between CGI density and $\log_{10}$ (chromosome size) (Figure 2(A)).

Also extremely significant positive correlation were discovered among CGI density with GC content (r=0.83, p=2×10^{-8}) (Figure 2(B)) and Obs_{CpG}/Exp_{CpG} (r=0.90, p=8.3×10^{-12}) (Figure 2(C)) on the chromosomes. These results indicate that chromosome GC content is probably a main genetic factor impacting CGI density. In addition CGI density was significantly positively correlated with gene density (r=0.74, p=2.8×10^{-6}) (Figure 2(D)). The pattern of significant correlation between CGI densities and some genomic features at the chromosome level such as chromosome size, GC content and Obs_{CpG}/Exp_{CpG} was similar to other studies (Han *et al.* 2008; Han and Zhao, 2008).

The significant positive correlation between CGI density and gene density in chromosome level was in agreement with research of Han and Zhao (2009) in dog genome (r=0.63, P=8.0×10^{-6}). Most of CGIs are sites of transcription initiation, including thousands that are remote from currently annotated promoters. Shared DNA sequence features adapt CGIs for promoter function by destabilizing nucleosomes and attracting proteins that create a transcriptionally permissive chromatin state.
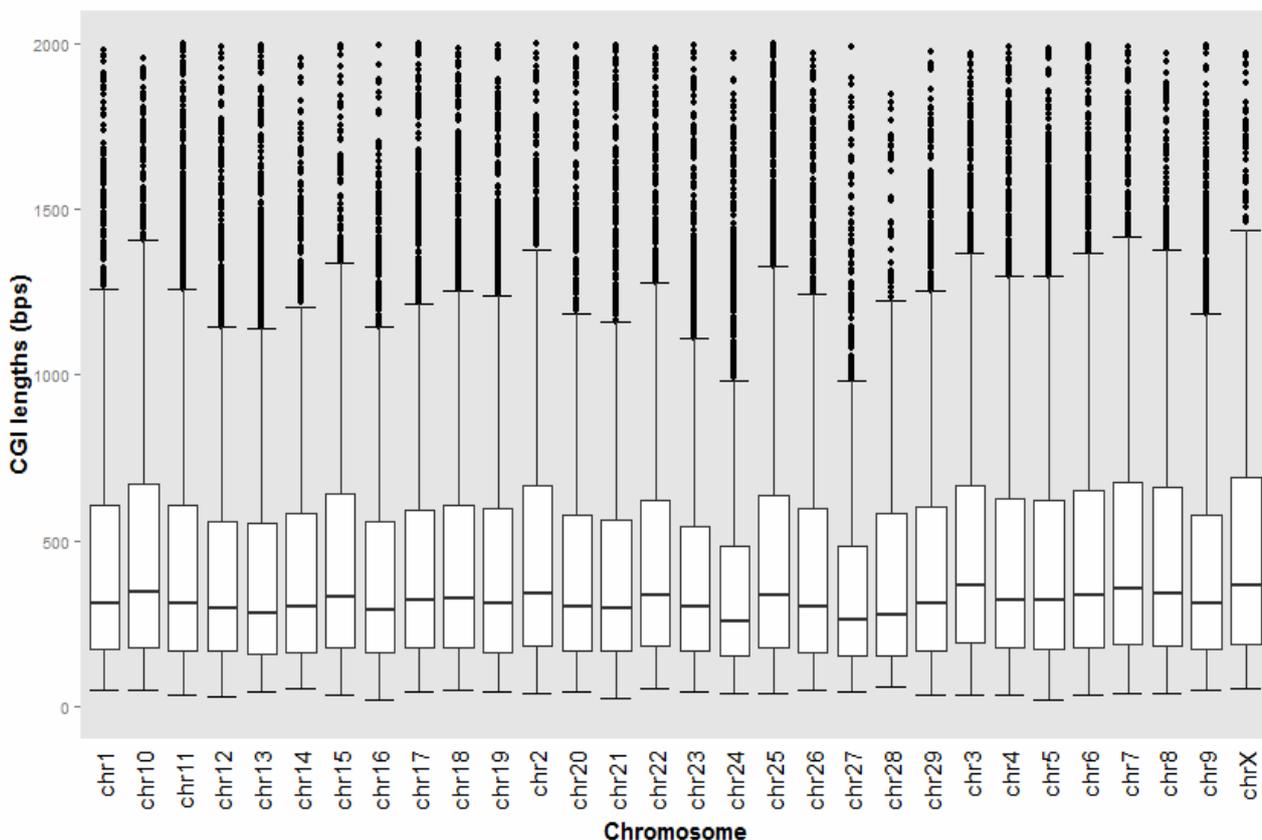

Figure 1 Lengths of CGIs in chromosomes of cattle by HMM algorithm
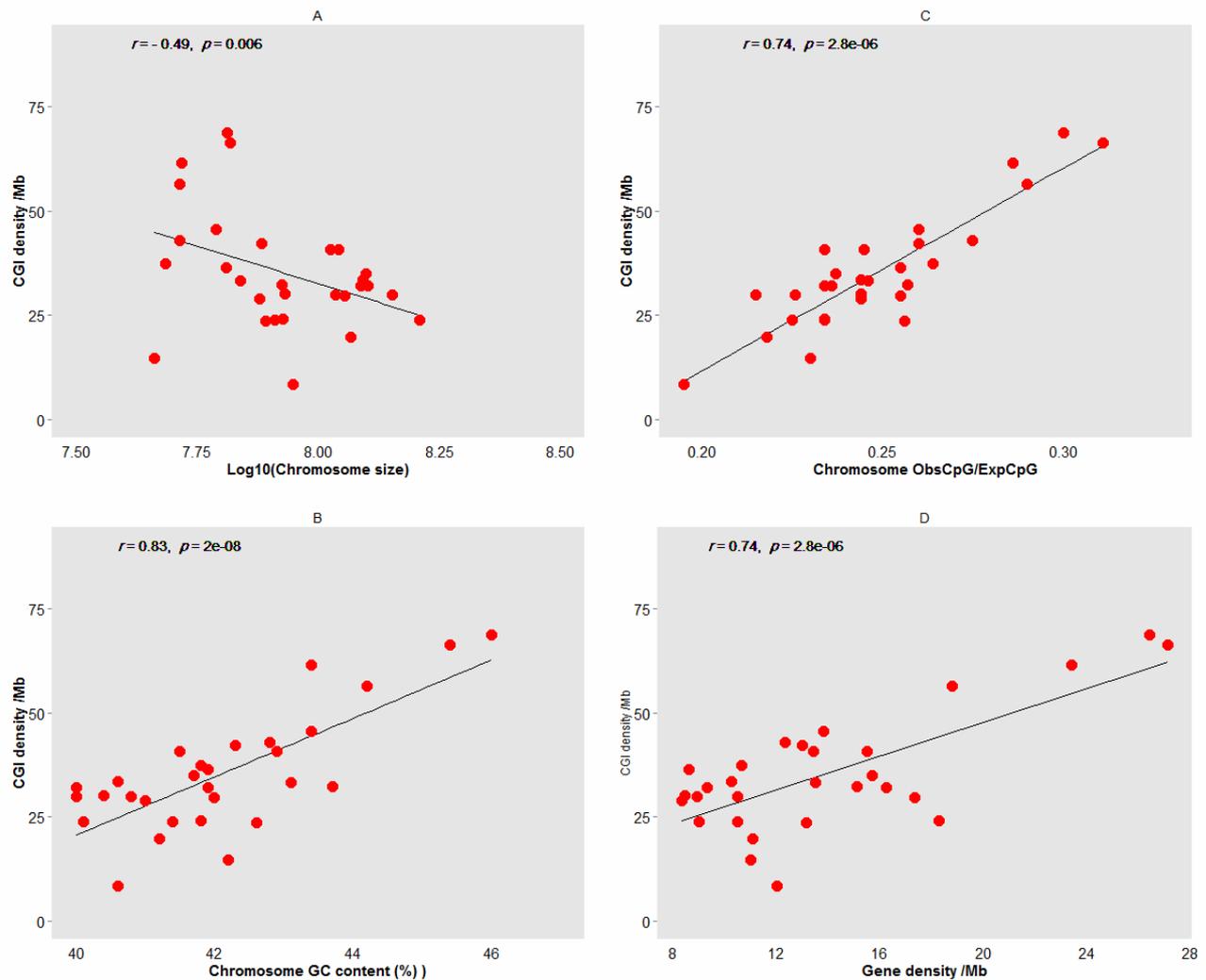Small difference in the average length and variance of CGI between chromosomes

Figure 2 Correlations between CGI densities with genomic features in cattle genome
(A) CGI density (per Mb) versus $\log_{10}$ (chromosome size)
(B) CGI density (per Mb) versus chromosome GC content (%)
(C) CGI density (per Mb) versus chromosome ObsCpG/ExpCpG
(D) CGI density (per Mb) versus Gene density (per Mb)
The Y chromosomes were excluded because of insufficient data

CGIs are therefore generically equipped to influence local chromatin structure and simplify regulation of gene activity (Deaton and Bird, 2011). Table 2 shows mean of CGI densities in segregated chromosomes by sizes into different groups (<50, 50-100, 100-150 and >150 Mb). When the size of chromosomes increased, the CGI densities decreased. The CGI density in the smallest chromosome group was about two times greater than the largest chromosome group. Previous study of GC content and CGIs in the chicken genomes (Hillier *et al.* 2004; Rao *et al.* 2013) and some mammalian genomes (Han *et al.* 2008) revealed a high density of CGIs on microchromosomes.

Interestingly, in the current study, the CGI density in the smallest chromosome group was greater than largest chromosome and it was consistent with those studies.

Table 2 CGIs densities in chromosomes with different sizes in cattle genome. The Y chromosomes were excluded because of insufficient data

Chromosome Size (Mb)	Number of chromosomes	Mean of CGI density/Mb±SD
< 50	3	52.74±47.67
50-100	15	41.06±15.56
100-150	11	30.10±9.23
> 150	1	23.90±0.0
Total	30	37.64±18.99

These results increase the possibility that gene density on microchromosomes approaches the maximum value known for vertebrates (McQueen *et al.* 1996). CGIs overlapped or within genes (gene-associated CGI) and CGIs in the intergenic regions (intergenic CGI) of cattle genome, correlated with genomic features.

An extremely significant correlation was discovered between gene-associated CGI densities with $\log_{10}$ (chromosome size) (r=-0.69, p=2.4×10^{-5}), with GC content of the chromosomes (r=0.90, p=1.1×10^{-11}) and Obs_{CpG}/Exp_{CpG} of the chromosomes (r=0.95, p=8.9×10^{-16}). However, correlation among intergenic CGI densities with $\log_{10}$ (chromosome size) (r=-0.39, p=0.03), GC content of the chromosomes (r=0.47, p=0.008) and Obs_{CpG}/Exp_{CpG} of the chromosomes (r=0.65, p=8.5×10^{-5}) were very low and also less significant than gene-associated CGI densities. These findings support the opinion that, CGIs can function as gene markers. The significant positive correlation between CGI density with GC content and gene density indicates that CGIs depend on both local genomic features and gene number (Han and Zhao, 2009). Also for a more detailed investigation, cattle gene annotation data was used to search for CGIs in different regions of genome.

Table 3 displays significant correlation among CGI densities in all genomic regions of cattle and genomic features ($\log_{10}$ (chromosome size), GC content and Obs_{CpG}/Exp_{CpG}) at the chromosome level. According to gene annotation in the NCBI and UCSC databases, 61095, 29613, 48759 and 12315 CGIs were overlapped with genes (gene–associated CGIs), intergenic regions (intergenic CGIs), intragenic regions (intragenic CGIs) and transcriptional start sites (TSS CGIs), respectively.

Number of CGIs and corresponding CGI density in intergenic region was remarkably lower than that in the intragenic and genes regions and had lower significant correlation with other genomic features (Table 3).

Other researcher founded similar results for CGIs in various genomic regions (Han *et al.* 2008; Han and Zhao, 2009; Medvedeva *et al.* 2010). These observations imply that CGIs are a considerable gene feature and they can be used to identify transcripts in cattle genomes.

CGI density and recombination rate in cattle genome
A set of recombination rate data of cattle (window size, 1 Mb was obtained from (Weng *et al.* 2014). A meaningful correlation (r=0.61, p=0.00031) was detected between CGI density and recombination rate (Figure 3(a)). Because the recombination rate increases from centromeric toward telomeric regions (Jensen-seaman *et al.* 2004; Han *et al.* 2008; Poissant *et al.* 2010; Weng *et al.* 2014), the trend of CGI density in length of chromosomes can be obtained. Attractively, a trend of higher CGI density in the telomeric regions is obtained (Figure 3(b)). This feature may be the reason of a positive correlation between CGI density and recombination rate. Several GC related measures (GC-rich repeats, CpG dinucleotide sites and CpG islands) were positively correlated with recombination rate in previous studies (Han *et al.* 2008; Tortereau *et al.* 2012; Rao *et al.* 2013).

Table 3 Correlation between CpG island (CGI) density in various cattle genomic regions and genomic features

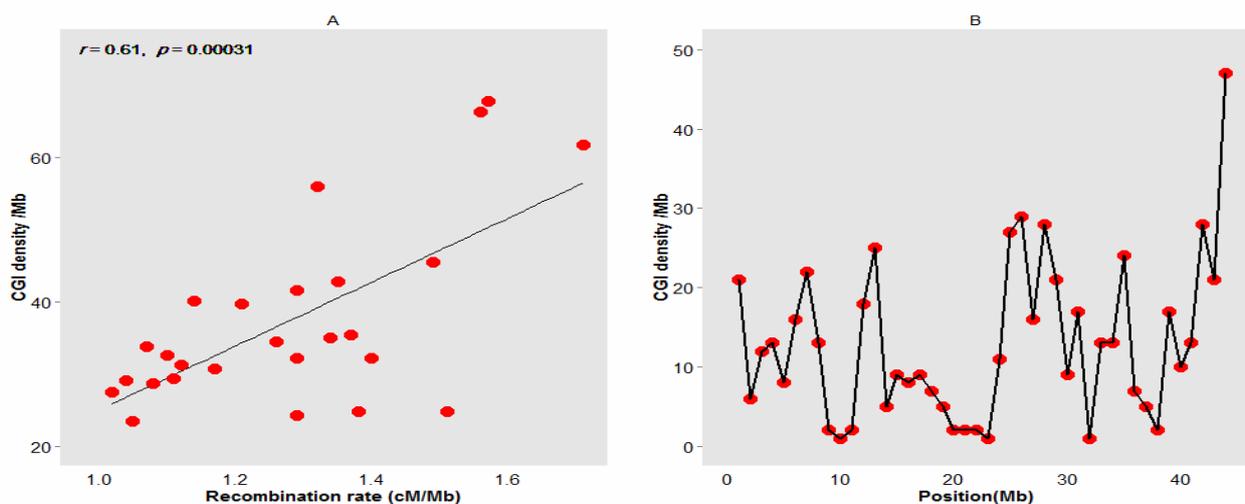| Items	Gene-associated CGIs (61095)		Intergenic CGIs (29613)		Intragenic CGIs (48795)		TSS CGIs (12315)	
	r	p	r	p	r	p	r	P
Log$_{10}$ (chromosome size)	-0.61	2.9×10^{-5}	-0.39	0.03	-0.48	0.01	-0.53	0.003
Guanine-cytosine content	0.91	4.1×10^{-12}	0.47	0.01	0.85	2.3×10^{-9}	0.87	3.2×10^{-10}
Obs$_{CpG}$/Exp$_{CpG}$	0.94	3.1×10^{-14}	0.65	8.5×10^{-5}	0.83	1.8×10^{-8}	0.86	8.6×10^{-10}

Figure 3 (A) correlation among CGI density and recombination rate (cM/Mb) in the cow genome,
(B) Distribution of CGI density (per Mb) on chromosome 28
The data demonstrate a trend of higher CGI density in telomeric regions
A similar trend was found for other chromosomes

The same result was observed in this study (Figure 3(a)). It is difficult to reveal the causality of the observed relationships between GC related measures such as CGIs and recombination rate, i.e., which parameter drives which. Further analyses of the mechanisms underlying recombination are needed to identify the molecular mechanism.

Comparison of CGIs in cattle genome and other vertebrate genomes

To detect the information on difference of CGI density between cattle and other vertebrate genomes, CGI density was scanned in eleven vertebrate genomes, including the sheep, goat, dromedary camel, bactrian camel, alpaca, human, mouse, horse, dog, chicken and zebrafish genomes (Figure 4).
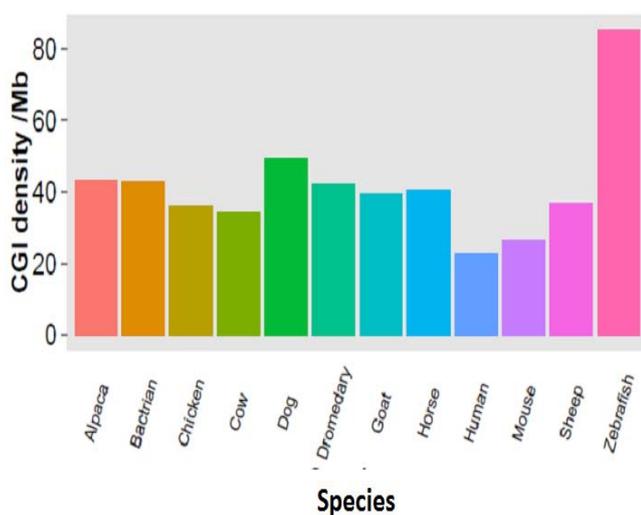

Figure 4 Cooperation of CGI density (per Mb) in various vertebrate genomes. CGI density varied among genomes

Predicted CGI density ranged from 23.05 (human) to 85.17 (zebrafish). Variation of CGI density between cattle with, chicken and other mammals (including ruminants) was reported to be not very high. According to previous studies, there are a lot of similarities between cattle genome and other ruminant as they have a close phylogeny, syntenic maps and orthologous genes (Elsik *et al.* 2009; Archibald *et al.* 2010; Jirimutu *et al.* 2012; Dong *et al.* 2013; Wu *et al.* 2014).

Probably, this is the reason of almost close CGIs density among these genomes. In mammals, the dog genome had the largest CGI density in comparison to other mammals and especially cattle. Han and Zhao (2009) studied contrast features of CpG islands in the promoter and other regions in the dog genome. They revealed a remarkably higher CGI density in the dog genome than in the human and mouse genomes.

But the dog genome had fewer promoter-associated CGIs than the human and mouse, and the abundance of CGIs in the dog genome was largely contributed by the non coding regions including the intergenic and intronic regions. The zebra fish genome had the largest CGI density in comparison to cattle and other ruminant genomes. CpGs are progressively depleted from fish to mammals.

This is mostly consistent with the idea of a greater degree of CpG loss with organism complexity (Irizarry *et al.* 2009). Variation in the number of CGIs and CGI density in warm-blooded vertebrate such as mammals, but very high in fish (Han *et al.* 2008). Fish genome is different in both number of CGIs and CGI density. In the study of Han and Zhao (2008) on fish genomes, they found that number of CGIs and the CGI density varied greatly in four fish genomes. They concluded that this feature might be caused by genetic (sequence composition evolution) and environmental factors such as water temperature, speed of flow, extent of light in different depth of water during the long evolutionary period after the divergence of common ancestor of fishes.

CONCLUSION

A systematic comparative genomic analysis of CGIs and CGI density was conducted in the cattle genome. The number of CGIs and the CGI density showed low variations in chromosomes. This study discloses significant correlations between CGI density and genomic features such as chromosome size, GC content, ObsCpG/ExpCpG, gene density and recombination rate in cattle. The results indicates neatly a close relationship between cattle evolution and CGI density.When comparing CGI between cattle and other vertebrates, it is evident that the former are characterized by lower number of CGI and CGI density when opposed to col-blooded vertebrates such as fishes. This is mostly consistent with the idea of a greater degree of CpG loss with organism complexity.

REFERENCES

Amiri Roudbar M., Mohammadabadi M.R. and Salmani V. (2015). Epigenetics: a new challenge in animal breeding. *Genet. Third Mel.* **12(4),** 3736-3751.

Archibald A.L., Cockett N.E., Dalrymple B.P., Faraut T., Kijas J.W., Maddox J.F., McEwan J.C., Hutton Oddy V., Raadsma H.W. and Wade C. (2010). The sheep genome reference sequence: a work in progress. *Anim. Genet.* **41,** 449-453.

Chuang L.Y., Huang H.C., Lin M.C. and Yang C.H. (2011). Particle swarm optimization with reinforcement learning for the prediction of CpG islands in the human genome. *PLoS One.* **6**, 21-36.

Chuang L.Y., Yang C.H. and Lin M.C. (2012). CpGPAP: CpG island predictor analysis platform. *BMC Genet*. **13,** 13.

Deaton A. and Bird A. (2011). CpG islands and the regulation of transcription. *Gen. Dev*. **25,** 1010-1022.

Dong Y., Xie M., Jiang Y., Xiao N., Du X., Zhang W., Tosser-Klopp G., Wang J., Yang S. and Liang J. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol*. **31,** 135-141.

Elsik C.G., Tellam R.L., Worley K.C., Gibbs R.A., Muzny D.M., Weinstock G.M., Adelson D.L., Eichler E.E., Elnitski L. and Guigó R. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*. **324,** 522-528.

Gardiner-Garden M. and Frommer M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol*. **196,** 261-282.

Glass J.L., Thompson R.F., Khulan B., Figueroa M.E., Olivier E.N., Oakley E.J., Zant G. Van Bouhassira E.E., Melnick A. and Golden A. (2007). CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids. Res*. **35,** 6798-6807.

Hackenberg M., Previti C., Luque-Escamilla P.L., Carpena P., Martinez-Aroza J. and Oliver J.L. (2006). CpGcluster: a distance-based algorithm for CpG island detection. *BMC Bioinformatics*. **7,** 446.

Hackenberg M., Barturen G., Carpena P., Luque-Escamilla P.L., Previti C. and Oliver J.L. (2010). Prediction of CpG island function: CpG clustering *vs*. sliding window methods. *BMC Genom*. **11,** 327.

Han L., Su B., Li W.H. and Zhao Z. (2008). CpG island density and its correlations with genomic features in mammalian genomes. *Genome. Biol*. **9(5),** 1-12.

Han L. and Zhao Z. (2008). Comparative analysis of CpG islands in four fish genomes. *Comp. Funct. Genomics*. **2008,** 565631.

Han L. and Zhao Z. (2009). Contrast features of CpG islands in the promoter and other regions in the dog genome. *Genomics*. **94,** 117-124.

Hillier L.W., Miller W., Birney E., Warren W., Hardison R.C., Ponting C.P., Bork P., Burt D.W., Groenen M.A.M. and Delany M.E. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. **432,** 695-716.

Irizarry R.A., Wu H. and Feinberg A.P. (2009). A species-generalized probabilistic model-based definition of CpG islands. *Mamm. Genome*. **20,** 674-680.

Jensen-Seaman M.I., Furey T.S., Payseur B.A., Lu Y., Roskin K.M., Chen C., Thomas M.A., Haussler D. and Jacob H.J. (2004). Comparative recombination rates in the rat, mouse, and human Genomes. *Genome. Res*. **14,** 528-538.

Jirimutu Wang Z., Ding G., Chen G., Sun Y., Sun Z., Zhang H., Wang L., Hasi S. and Zhang Y. (2012). Genome sequences of wild and domestic bactrian camels. *Nat. Commun*. **3,** 1202-1212.

McQueen H.A., Fantes J., Cross S.H., Clark V.H., Archibald A.L. and Bird A.P. (1996). CpG islands of chicken are concentrated on microchromosomes. *Nat. Genet*. **12,** 321-324.

Medvedeva Y.A., Fridman M.V, Oparina N.J., Malko D.B., Ermakova E.O., Kulakovskiy I.V, Heinzel A. and Makeev V.J.

(2010). Intergenic, gene terminal and intragenic CpG islands in the human genome. *BMC Genomics*. **11,** 48.

Poissant J., Hogg J.T., Davis C.S., Miller J.M., Maddox J.F. and Coltman D.W. (2010). Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics*. **11,** 524.

Quinlan A.R. and Hall I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26,** 841-842.

Rabiner L.R. (1989). Tutorial on hmm and applications. *IEEE J*. **77,** 257-286.

Rao Y.S., Chai X.W., Wang Z.F., Nie Q.H. and Zhang X.Q. (2013). Impact of GC content on gene expression pattern in chicken. *Genet. Sel. Evol*. **45,** 9.

Rice P., Longden I. and Bleasby A. (2000). EMBOSS: the European molecular biology open software suite. *Trends. Genet*. **16,** 276-277.

Saito Y., Nagae G., Motoi N., Miyauchi E., Ninomiya H., Uehara H., Mun M., Okumura S., Ohyanagi F. and Nishio M. (2016). Prognostic significance of CpG island methylator phenotype in surgically resected small cell lung carcinoma. *Cancer Sci*. **107(3),** 320-325.

Su J., Zhang Y., Lv J., Liu H., Tang X., Wang F., Qi Y., Feng Y., and Li X. (2010). CpG_MI: a novel approach for identifying functional CpG islands in mammalian genomes. *Nucleic Acids. Res*. **38,** 6.

Takai D. and Jones P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA*. **99,** 3740-3745.

Tan B., Yin Y., Liu Z., Tang W., Xu H., Kong X., Li X., Yao K., Gu W. and Smith S.B. (2011). Dietary L-arginine supplementation differentially regulates expression of lipid-metabolic genes in porcine adipose tissue and skeletal muscle. *J. Nutr. Biochem*. **22,** 441-445.

Tortereau F., Servin B., Frantz L., Megens H.J., Milan D., Rohrer G., Wiedmann R., Beever J., Archibald A.L. and Schook L.B. (2012). A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*. **13,** 586.

Vercelli D. (2016). Does epigenetics play a role in human asthma? *Allergol. Int*. **65,** 123-126.

Wang Y. and Leung F.C.C. (2004). An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics*. **20,** 1170-1177.

Weng Z.Q., Saatchi M., Schnabel R.D., Taylor J.F. and Garrick D.J. (2014). Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genet. Sel. Evol*. **46,** 34-41.

Willham R.L. (1986). From husbandry to science: a highly significant facet of our livestock heritage. *J. Anim. Sci*. **62,** 1742-1758.

Wu H., Caffo B., Jaffee H.A., Irizarry R.A. and Feinberg A.P. (2010). Redefining CpG islands using hidden Markov models. *Biostatistics*. **11,** 499-514.

Wu H., Guang X., Al-Fageeh M.B., Cao J., Pan S., Zhou H., Zhang L., Abutarboush M.H., Xing Y. and Xie Z. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun*. **5,** 5188-5197.

Yang C.H., Lin Y.D., Chiang Y.C. and Chuang L.Y. (2016). A hybrid approach for CpG island detection in the human genome. *PLoS One*. **11,** 23-36.