

# Comparative analysis of sequence characteristics of imprinted genes in human, mouse, and cattle

Hasan Khatib · Ismail Zaitoun · Eui-Soo Kim

Received: 28 February 2007 / Accepted: 23 May 2007 / Published online: 26 July 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Genomic imprinting is an epigenetic mechanism that results in monoallelic expression of genes depending on parent-of-origin of the allele. Although the conservation of genomic imprinting among mammalian species has been widely reported for many genes, there is accumulating evidence that some genes escape this conservation. Most known imprinted genes have been identified in the mouse and human, with few imprinted genes reported in cattle. Comparative analysis of genomic imprinting across mammalian species would provide a powerful tool for elucidating the mechanisms regulating the unique expression of imprinted genes. In this study we analyzed the imprinting of 22 genes in human, mouse, and cattle and found that in only 11 was imprinting conserved across the three species. In addition, we analyzed the occurrence of the sequence elements CpG islands, C + G content, tandem repeats, and retrotransposable elements in imprinted and in nonimprinted (control) cattle genes. We found that imprinted genes have a higher G + C content and more CpG islands and tandem repeats. Short interspersed nuclear elements (SINEs) were notably fewer in number in imprinted cattle genes compared to control genes, which is

in agreement with previous reports for human and mouse imprinted regions. Long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs) were found to be significantly underrepresented in imprinted genes compared to control genes, contrary to reports on human and mouse. Of considerable significance was the finding of highly conserved tandem repeats in nine of the genes imprinted in all three species.

## Introduction

Genomic imprinting is an epigenetic mechanism that results in monoallelic expression of genes depending on parent-of-origin of the allele. Several studies have shown that genomic imprinting is conserved among mammalian species. However, there is accumulating evidence of the lack of conservation of imprinted genes. In a survey of 63 examined protein-coding imprinted genes, 35 in human and 54 in mouse, only 26 genes were reported to be conserved in both species (Morison et al. 2005). Paulsen et al. (2000) investigated the sequence conservation among six genes on distal chromosome 7 in the mouse and their human orthologs on chromosome 11p15. Although the organization of the mouse and human genes was found to be highly conserved, some genes showed nonimprinting patterns in both species.

Most known imprinted genes have been identified in the mouse and human, and few genes have been reported to be imprinted in cattle. Recently, however, we and others have reported the conserved imprinting of several bovine genes, including *IGF2R* (Killian et al. 2001); *NESP55* (Khatib 2004); *IGF2*, *MEG3*, *XIST* (Dindot et al. 2004); *PEG3* (Kim et al. 2004); *H19*, *NAPIL5*, and *NNAT* (Zaitoun and

---

**Electronic supplementary material** The online version of this article (doi: 10.1007/s00335-007-9039-z) contains supplementary material, which is available to authorized users.

---

H. Khatib (✉) · I. Zaitoun  
Department of Dairy Science, University of Wisconsin-Madison, 1675 Observatory Dr., Madison, Wisconsin 53706, USA  
e-mail: hkhatib@wisc.edu

E.-S. Kim  
Department of Animal Sciences, University of Wisconsin-Madison, 1675 Observatory Dr., Madison, Wisconsin 53706, USA

Khatib 2006). Conversely, we reported that imprinting of *DCN*, *SDHD*, *COPG2* (Khatib 2005a), and *SLC38A4* (Zaitoun and Khatib 2006) is not conserved among human, mouse, and cattle. Comparative analysis of genomic imprinting across mammalian species would offer a powerful tool for elucidating the mechanisms regulating the unique expression of imprinted genes (Dindot et al. 2004; Killian et al. 2001). Hence, the intent of this study is to compare the imprinting status of cattle genes with that of the human and mouse orthologs.

Recently, the question of whether imprinted genes have sequence characteristics that distinguish them from non-imprinted genes is drawing the attention of several research groups. Such structural differences might elucidate the mechanisms leading to allele-specific expression of imprinted genes (Okamura and Ito 2006). Greally (2002) found that the main sequence characteristic of human imprinted genes is a lower incidence of short interspersed nuclear elements (SINEs). Allen et al. (2003) reported that the densities of CpG islands and SINEs were lower in the flanking regions of monoallelically expressed and imprinted genes when compared to biallelically expressed genes. They also showed that monoallelically expressed and imprinted genes were flanked by regions with a high density of long interspersed nuclear elements (LINEs) compared to biallelically expressed genes (Allen et al. 2003). In a search for sequence features in the *IGF2* gene in mammalian species where it is either imprinted or nonimprinted, Weidman et al. (2004) found that the imprinting of *IGF2* is strongly associated with a lack of SINEs. Shirohzu et al. (2004) found that a 210-kb DNA segment, located on mouse chromosome 7 between two imprinting loci, was rich in tandem repeats, LINE-1 elements, and retroviral insertions. Additional support for the hypothesis of the lower density of SINEs defining imprinted regions was provided by Walter et al. (2006). They reported that the frequency of SINEs in imprinted genes was 7.32% compared to 13.9% in nonimprinted genes. Also, they reported that in genes with a G + C content of greater than 40%, the frequency of LINEs was higher in imprinted genes than in control genes. An interesting structural element that distinguishes imprinted from non-imprinted genes was reported for the *IMPACT* gene, which is known to be imprinted in mouse but not in human. The imprinting status of *IMPACT* was found to be associated with the presence of tandem repeats in the CpG islands of the mouse gene, whereas the nonimprinted gene in human lacks repeats and differential methylation in CpG islands (Okamura and Ito 2006). A recent study that included 38 imprinted human, 39 imprinted mouse, and 79 control genes showed that CpG islands of imprinted genes were enriched in tandem repeats compared to nonimprinted genes (Hutter et al. 2006).

Giving structural elements a role in the regulation and maintenance of genomic imprinting is an attractive hypothesis, and as these elements have not yet been reported in bovine sequences, the objectives of this study were to evaluate the occurrence of G + C content, CpG islands, and retrotransposable elements in bovine genes and to perform comparative analysis between cattle imprinted and nonimprinted genes and among cattle and mouse and human genes. Such inter- and intraspecies comparative analyses would provide insight into the mechanisms and the evolution of genomic imprinting.

## Methods

### Gene selection

A total of 22 genes with known imprinting status in cattle were selected for comparative analysis; *IGF2*, *IGF2R*, *MEG3*, *MEST*, *PEG3*, *XIST*, and *ZIM2* were selected from the literature (Table 1), whereas *DCN*, *NESP55-GNAS*, *H19*, *MAGEL2*, *NAP1L5*, *NNAT*, *RTL1*, *SDHD*, *SLC38A4*, *ASB4*, *CD81*, *HTR2A*, *OSBPL5*, *PEG10*, and *TSSC4* were selected based on work in our laboratory. Data on the imprinting status of the orthologous human and mouse genes were obtained from the literature and from the Catalogue of Imprinted Genes (<http://www.igc.otago.ac.nz/>). For comparison with the imprinted genes, 20 genes that had previously been determined to show biallelic expression in various experiments in our laboratory were selected to serve as controls (Table 2).

### Sequence analysis

Imprinted and control gene sequences were studied in two ways. In the first analysis, a total of 245,823 bp of coding sequence from 14 imprinted cattle genes (Table 2) and a total of 909,448 bp of coding sequence from the 20 control genes were searched and CpG islands, direct tandem repeats, SINEs, LINEs, long terminal repeats (LTRs), and G + C content were quantified. Also, a total of 8,821,311 bp from the bovine X chromosome were searched for LINE elements. In the second analysis, 50 kb of upstream sequence from the transcription initiation site and 50 kb of downstream sequence from the transcription termination site, in addition to the coding regions of the same 34 genes, were analyzed for the aforementioned sequence elements. The exact sizes of the upstream and downstream regions of each gene were determined considering the location of potential regulatory regions. *TCCS4* was not included in the second analysis because its flanking sequences have not yet been identified. For this analysis, a total of 2571 genes from bovine chromosome 5 ( $n = 1335$ ), chromosome 6 ( $n =$

**Table 1** Comparative imprinting analysis of human, mouse, and cattle genes

Gene	Human <sup>a</sup>	Mouse <sup>a</sup>	Cattle	Expression	Reference for cattle
<i>DCN</i>	NI <sup>b</sup>	I	NI	M	Khatib 2005a
<i>NESP55-GNAS</i>	I	I	I	M	Khatib 2004
<i>H19</i>	I	I	I	M	Zaitoun and Khatib 2006; Zhang et al. 2004
<i>IGF2</i>	I	I	I	P	Dindot et al. 2004
<i>IGF2R</i>	PI	I	I	M	Killian et al. 2001
<i>MAGEL2</i>	I	I	I	P	Unpublished data
<i>MEG3</i>	I	I	I	M	Dindot et al. 2004
<i>MEST</i>	I	I	I	P	Ruddock et al. 2004
<i>NAP1L5</i>	I <sup>c</sup>	I	I	P	Zaitoun and Khatib 2006
<i>NNAT</i>	I	I	I	P	Zaitoun and Khatib 2006
<i>PEG3</i>	I	I	I	P	Kim et al. 2004
<i>RTL1</i>	I	I	I	P	Unpublished data
<i>SDHD</i>	CD	NR	NI	P	Khatib 2005a
<i>SLC38A4</i>	NR	I	NI	P	Zaitoun and Khatib 2006
<i>XIST</i>	NI	I	I	P	Dindot et al. 2004
<i>ZIM2</i>	I	I	NI	M, P	Kim et al. 2004
<i>ASB4</i>	NR	I	NI	M	Unpublished data
<i>CD81</i>	NI	I	NI	M	Unpublished data
<i>HTR2A</i>	CD	I	NI	M	Unpublished data
<i>OSBPL5</i>	I	I	NI	M	Unpublished data
<i>PEG10</i>	I	I	I	P	Unpublished data
<i>TSSC4</i>	NI	I	I	M	Unpublished data

NR = not reported; I = imprinted; PI = polymorphic imprinting; NI = not imprinted; M = maternal; P = paternal; CD = conflicting data

<sup>a</sup> Data from <http://www.otago.ac.nz/IGC>

<sup>b</sup> Monk et al. 2006

<sup>c</sup> Wood et al. 2007

565), and chromosome 14 ( $n = 671$ ) were selected to serve as a controls for SINE and LINE densities.

CpG islands were identified using the EMBOSS program *cpplot* (<http://www.bioweb.pasteur.fr/seqanal/interfaces/cpgplot.html>) with a window size of 120 bp, window shift increment of 1 bp, minimum length of an island of 200 bp minimum observed-to-expected ratio of CpG dinucleotides of 0.6, and minimum percentage of 50 (C + G content) (Gardiner-Garden and Frommer 1987). C + G content was calculated using the EMBOSS program *geecee* (<http://www.bioweb.pasteur.fr/docs/EMBOSS/geecee.html>) or RepeatMasker (<http://www.repeatmasker.org>). Direct tandem repeats were identified using the Tandem Repeats Finder version 3.21 (<http://www.tandem.bu.edu/trf/trf.html>). The alignment parameters for match, mismatch, and indels were 2, 5, and 7, respectively. The minimum alignment score to report repeat was 100 and the maximum period size was 2000 bp.

The multiple sequence alignment program ClustalW (<http://www.ebi.ac.uk/clustalw>) was used to align sequences of direct tandem repeats identified in imprinted genes in human, mouse, and cattle. Repeat masking was

done using RepeatMasker open version 3.1.6 (<http://www.repeatmasker.org>). This program searches DNA sequences in FASTA format for interspersed repeats and returns a masked query sequence ready for database searches. Perl script was used to calculate transposable element frequency for each gene based on RepeatMasker results. To test for statistically significant differences between imprinted and nonimprinted genes regarding their structural sequence elements (C + G content, SINEs, LINEs, and LTRs), we used the Wilcoxon-Mann-Whitney test. Boxplots of the distributions of LINE-1 and SINE frequencies were plotted using the statistics package R (R-project, <http://www.cran.r-project.org/>).

## Results

Conservation of genomic imprinting in mammalian species

Table 1 shows the imprinting status of cattle genes and their human and mouse orthologs. Eleven genes were

**Table 2** The number of tandem repeats (TR) and CpG islands in coding sequences and in total sequences (coding and 50-kb up- and downstream flanking sequences) of imprinted and control genes

Gene	Coding sequence (bp)	TR in coding sequences	TR in total sequences	CpG islands in coding sequences	CpG islands in total sequences
<i>NESP55-GNAS</i>	13523	1	4	0	2
<i>IGF2</i>	6070	0	22	2	19
<i>IGF2R</i>	102613	11	23	31	36
<i>MAGEL2</i>	4000	4	5	1	4
<i>MEST</i>	13919	1	5	1	3
<i>NAP1L5</i>	2354	0	1	1	1
<i>NNAT</i>	1609	0	1	2	4
<i>PEG3</i>	23509	11	18	19	30
<i>RTL1</i>	3996	0	29	4	10
<i>XIST</i>	36421	6	19	0	0
<i>PEG10</i>	6978	1	5	1	5
<i>TSSC4</i>	3661	0	NA	1	NA
<i>H19</i>	1305	0	19	3	22
<i>MEG3</i>	25865	2	22	4	15
Total imprinted	218653	37	173	70	151
<i>ARF1</i>	18062	0	10	3	15
<i>GHR</i>	173709	5	12	0	0
<i>HP</i>	6161	0	3	1	8
<i>OLR1</i>	11306	0	4	0	0
<i>PLP1</i>	15185	0	8	0	2
<i>TKT1</i>	13498	0	1	0	5
<i>WARS</i>	22458	1	22	2	22
<i>COL1A2</i>	36670	1	3	0	1
<i>GRP58</i>	22203	0	1	1	9
<i>JDP1</i>	34676	1	7	0	2
<i>PAFAH1B2</i>	26804	1	2	1	2
<i>TG</i>	236368	1	12	5	5
<i>UTMP</i>	8639	1	9	0	0
<i>DCN</i>	39226	1	9	0	0
<i>SLC38A4</i>	49311	1	4	0	3
<i>ZIM2</i>	16039	0	3	1	7
<i>ASB4</i>	78511	2	6	1	3
<i>CD81</i>	7071	0	3	0	10
<i>HTR2A</i>	61202	0	6	0	2
<i>OSBPL5</i>	32344	3	22	6	28
Total control	909443	19	150	21	124

found to be imprinted in all three species; of those, eight genes were paternally expressed and three were maternally expressed. Of 22 genes known to be imprinted in human or in mouse, 14 were found to be imprinted in cattle. Of the remaining genes, *DCN*, *CD81*, *SLC38A4*, and *ASB4*—known to be imprinted in mouse—were found to be not imprinted in cattle. For *SDHD* and *HTR2A*, conflicting data have been reported regarding their imprinting status in human. The *ZIM2* and *OSBPL5* genes were reported to be

imprinted in human and mouse but found to be not imprinted in cattle.

Sequence characteristics of coding sequences of imprinted and control genes

Genomic sequence characteristics of a total of 245,823 bp from the coding sequences of 14 imprinted cattle genes were compared to those of 909,443 bp of coding sequences

from 20 control genes (Table 2). The average number of tandem repeats per kb in imprinted genes was 0.15, whereas in control genes the average was 0.021 tandem repeats per kb. Similarly, the average number of CpG islands per kb was higher in imprinted (0.285) than in control genes (0.023). The frequency of G + C was also higher in imprinted genes (51%) than in control genes (45%) ( $p = 0.018$ ). For imprinted genes, 11 tandem repeats were found in CpG islands of three genes (*IGF2*, *MAGEL2*, and *PEG3*) compared to one tandem repeat found in one control gene (*WARS*). Figure 1 shows boxplots of the frequency of SINEs and LINE-1s in imprinted and control genes. The frequency of SINEs was significantly lower in imprinted genes, 5.5%, whereas the frequency of these elements in control genes was 16.2% ( $p < 0.0001$ ). The genes *NESP55*, *IGF2*, *MAGEL2*, *NNAT*, *RTL1*, and *PEG10* had no SINEs in their intragenic sequences, whereas *IGF2R*, *MEST*, *NAPIL5*, *PEG3*, *XIST*, and *TSSC4* had SINEs at frequencies that ranged from 2.1% to 10.7%. In contrast, SINEs were found in all of the 20 control genes examined in this study. The frequencies of SINEs in control genes ranged from 0.7% to 24.6%. Likewise, the frequency of LINES was significantly lower ( $p = 0.0003$ ) in imprinted genes (4.7%) than in control genes (13.7%). In contrast, frequency of LINES in a total of 8,821,311 bp from the bovine X chromosome was 26.8%.

Also, the frequency of LTRs was significantly lower ( $p = 0.0116$ ) in imprinted genes (0.4%) compared to control genes (1.7%). Only two imprinted genes (*IGF2R* and *XIST*) had these elements, whereas 11 of 20 control genes had LTRs with frequencies that ranged from 0.3% to 3.8%.

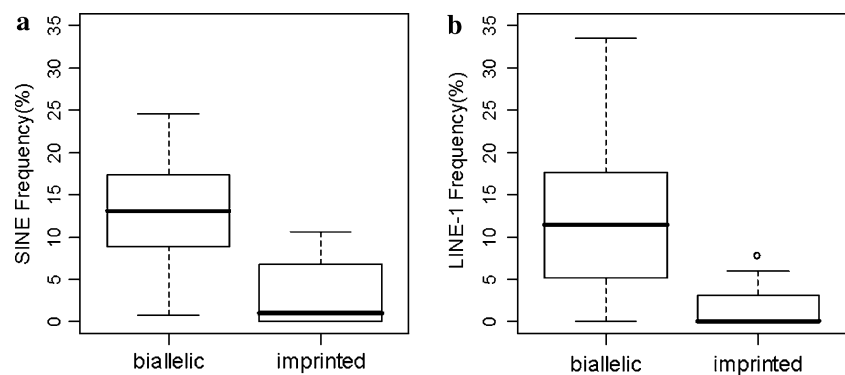
To test whether tandem repeats found in imprinted genes are conserved among mammalian species, we used the Tandem Repeats Finder to identify tandem repeats in the imprinted genes listed in Table 1. A total of 64, 50, and 45 tandem repeats were found in 12 human, 11 mouse, and 13 cattle genes, respectively. Supplementary Tables 1, 2, and 3 show repeat size, copy number, and sequence of tandem repeats found in human, mouse, and cattle genes, respectively. To identify homologous tandem repeats in human,

mouse, and cattle, we used the multiple sequence alignment program ClustalW (<http://www.ebi.ac.uk/clustalw>). Alignment was performed only for genes that were found to be imprinted in at least two species (see Table 1). Table 3 shows the number of tandem repeats found in nine genes and the sequence alignment of these repeats across human, mouse, and cattle species. For *GNAS*, a total of four alignments were found between human and mouse repeats, in which a 75% sequence similarity (score) was found. It is worth noting that H3 and M1 repeats that showed a score of 75% are located in CpG islands of *GNAS*. For *IGF2R*, six alignments were found between human/cow and human/mouse repeats. The highest alignment score (76%) was between human repeat 5 (H5) and cow repeat 1 (C1). The repeats H5, C1, C3, and C11 were found in CpG islands of *IGF2R*. *PEG3* showed 12 sequence alignments of tandem repeats in human, mouse, and cattle with scores that ranged from 53% to 90%: five human/cow alignments, four human/mouse alignments, and three alignments of mouse/cow (Table 3). Highly conserved tandem repeats across species with scores of at least 80% were found for *MAGEL2*, *MEST*, *PEG3*, *RTL1*, *OSBPL5*, and *PEG10*.

#### Sequence characteristics of flanking sequences of imprinted and control genes

For the analysis of flanking sequences of imprinted and control genes we chose a 50-kb window, considering the location of potential regulatory regions. Tandem repeats were found in all imprinted and control genes examined (Table 2). CpG islands were identified in all imprinted genes except for *XIST* and in 16 of 20 control genes (Table 2). Table 4 shows the average of number of tandem repeats and CpG islands per kb in coding and flanking sequences of imprinted and control genes. Of considerable interest was the high density of tandem repeats and CpG islands observed in coding sequences of imprinted genes (0.151/kb and 0.285/kb, respectively) vs. control genes (0.021/kb and 0.023/kb, respectively). Figure 2 and Supplementary Figure 1 show the distribution of CpG islands

**Fig. 1** Boxplot graphs (median and interquartile) representing (a) SINE and (b) LINE-1 frequencies identified by the RepeatMasker program in 14 imprinted cattle genes and in the coding sequences of 20 control (nonimprinted) cattle genes. Wilcoxon-Mann-Whitney test revealed a statistically significant difference for the occurrence of SINEs and LINE-1s between imprinted and control genes



**Table 3** Sequence alignment of tandem repeats in imprinted genes across human, mouse, and cattle

Gene	Number of tandem repeats			Aligned repeats <sup>a</sup>	Score	Repeats in CpG islands
	Human	Mouse	Cattle			
<i>GNAS</i>	7	13	1	H3/M1	75	H1, H2, H3, M1, M2
				H5/M7	75	
				H3/M2	72	
				H2/M9	64	
<i>IGF2R</i>	20	3	11	H5/C1	76	H5, C1, C3, C10, C11
				H5/M2	64	
				H5/C3	64	
				H5/C11	64	
				H20/C1	59	
				H20/C3	56	
<i>MAGEL2</i>	2	3	4	H2/C4	91	C2
				H1/C4	88	
				M2/C1	76	
				H1/C3	61	
				H2/M1	51	
<i>MEST</i>	2	0	1	H2/C1	89	
<i>PEG3</i>	7	7	11	H7/M6	90	
				H5/C9	80	
				H6/M5	80	
				M5/C9	76	
				H2/C3	72	
				H7/M2	66	
				H7/C1	66	
				M7/C10	66	
				H6/M3	60	
				H7/C9	57	
				M7/C9	57	
				H3/C2	53	
<i>RTL1</i>	2	2	0	H1/M1	80	
<i>XIST</i>	4	3	6	H2/C2	71	
				H4/M1	47	
<i>OSBPL5</i>	6	2	3	M2/H1	83	
				M2/H4	83	
				M2/H5	83	
				M2/C3	83	
<i>PEG10</i>	2	9	1	H2/M3	89	
				H2/M4	81	
				M1/C1	55	

<sup>a</sup> H = human, M = mouse, C = cattle

in coding and flanking sequences of imprinted genes compared with control genes. A higher density of CpG islands was observed in imprinted gene regions compared with biallelically expressed genes.

To test whether SINE and LINE-1 densities are different between coding sequences and flanking sequences, we calculated the frequencies of these elements in imprinted

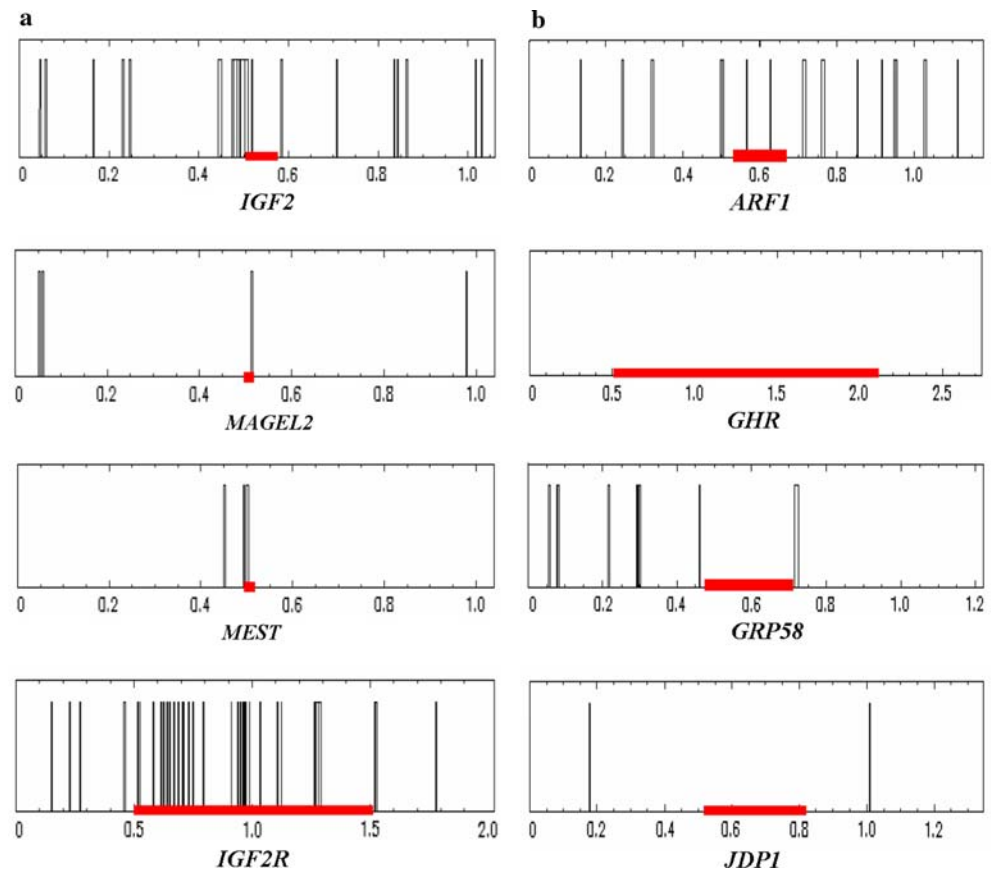
and control genes and in a total of 2571 genes from bovine chromosomes 5, 6, and 14. Figure 3 shows boxplots of SINE and LINE-1 frequencies in imprinted and control genes. The Wilcoxon-Mann-Whitney test revealed significantly lower SINE density in imprinted genes compared to biallelically expressed genes ( $p = 0.002$ ) and chromosome 14 ( $p = 0.0015$ ), chromosome 5 ( $p = 0.0005$ ), and chro-



**Table 4** Number of tandem repeats (TR) and CpG islands per kb genomic sequences in coding and flanking sequences of imprinted and biallelically expressed genes

Element/gene type	50-kb upstream	50-kb downstream	Coding sequence	Total sequence
TR/kb imprinted	0.120	0.100	0.151	0.113
TR/kb biallelic	0.071	0.060	0.021	0.052
CpG/kb imprinted	0.073	0.055	0.285	0.100
CpG/kb biallelic	0.054	0.048	0.023	0.043

**Fig. 2** Distribution of CpG islands in a subset of (a) imprinted and (b) biallelically expressed genes with 50-kb upstream and downstream flanking sequences. Vertical bars indicate position of each CpG island and the thick line on the x axis corresponds to the coding region of the gene. The position of each genomic region counts from the upstream flanking region of each gene ( $\times 10^5$  bp). The distribution of CpG islands in other genes examined is presented in Supplementary Fig. 1



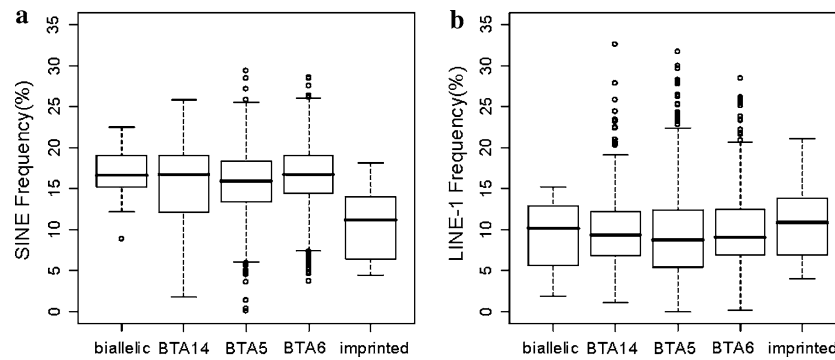
mosome 6 ( $p = 0.0001$ ) genes (Fig. 3a). In contrast, LINE-1 frequency was not significantly different between gene groups examined (Fig. 3b).

## Discussion

To understand the evolution of genomic imprinting and the mechanisms controlling allele-specific expression of imprinted genes, it is crucial to identify species-specific imprinted genes and compare their structural features and to identify sequence elements that differentiate imprinted from nonimprinted genes (Okamura and Ito 2006). In this study we investigated the conservation of imprinting of 22 genes in human, mouse, and cattle. In addition, we analyzed the occurrence of the sequence elements CpG

islands, C + G content, tandem repeats, and retrotransposable elements in imprinted and control cattle genes. Also, we investigated the conservation of tandem repeats located in imprinted genes in human, mouse, and cattle.

There is accumulating evidence of limited conservation of imprinted genes across species. Recently, we reported that the *SDHD* and *COPG2* genes were not imprinted in cattle and sheep tissues (Khatib 2005a, b). Also, Kim et al. (2004) showed that *ZIM2* was biallelically expressed in cattle in contrast to the monoallelic expression observed in human and mouse. Morison et al. (2005) reported that of 63 protein-coding imprinted genes, only 26 were imprinted in both human and mouse. Monk et al. (2006) reported the lack of imprinting of six human genes and the polymorphic imprinting of another three genes, all known to be placenta-specific, imprinted in the mouse.



**Fig. 3** Boxplots of retrotransposable element densities in imprinted genes, in biallelically expressed control genes, and in genes from bovine chromosome 14 ( $n = 671$ ), chromosome 5 ( $n = 1335$ ), and chromosome 6 ( $n = 565$ ). SINE and LINE-1 frequencies were calculated for the coding sequence and 50-kb upstream and downstream of the coding region. a SINE frequency in imprinted

genes was 10.89% compared with 16.57%, 15.42%, 15.65%, and 16.46% in biallelically expressed control genes and genes from BTA14, BTA5, and BTA6, respectively. b LINE-1 frequencies were 10.71%, 9.30%, 9.80%, 9.34%, and 9.93% in imprinted, biallelically expressed control genes, and genes from BTA14, BTA5, and BTA6, respectively

Recently, we investigated the imprinting status of *CD81*, *TSSC4*, and *OSBPL5*—reported to be placenta-specific imprinted in mouse—on bovine chromosome 29 and the cluster of *PEG10* and *ASB4* on bovine chromosome 4 (unpublished data). *CD81*, *TSSC4*, and *OSBPL5* were found to be expressed in all fetal tissues examined, including ovary, skeletal muscle, liver, pituitary, mammary gland, kidney, brain, spleen, heart, pancreas, eye, and caruncle. Imprinting analysis revealed biallelic expression of *CD81* in all cattle tissues examined, like the human gene (Monk et al. 2006), but in contrast to the mouse ortholog which was reported to be maternally expressed in placental tissues (Lewis et al. 2004). Species-specific imprinting was also observed for *OSBPL5*, which is known to be imprinted in mouse (Engemann et al. 2000) and human placenta (Higashimoto et al. 2000) but biallelically expressed in other tissues. In contrast to human and mouse, our study revealed biallelic expression of *OSBPL5* in both placental and nonplacental tissues (data not shown). Similarly, *TSSC4* was shown to be imprinted in cattle (data not shown) and mouse but not in human (Monk et al. 2006).

The number of known imprinted genes in cattle is small and the mechanisms regulating imprinting in this species are poorly understood. However, the lack of conservation of placenta-specific imprinted genes between mouse and human might be due to allele-specific histone modifications present in mouse but absent in human genes (Monk et al. 2006). This may be a contributing factor to why, for the placenta-specific genes reported in this study, the imprinting pattern was not conserved across human, mouse, and cattle. For *ASB4*, the imprinting status in human is not known and the mouse gene has been reported to be maternally expressed in a wide range of fetal tissues (Mizuno et al. 2002). In our study, although *ASB4*'s expression pattern was similar to that of mouse, the bovine

gene showed biallelic expression in all examined fetal tissues (data not shown). Thus, further studies involving more imprinted genes and more species may be necessary to confirm whether species-specific imprinting is the rule or the exception.

Analysis of characteristic sequence elements revealed that G + C content was significantly higher in imprinted cattle genes when compared to that of control genes. In contrast, Hutter et al. (2006) did not find significant differences in G + C content between imprinted genes and control sequences in human and mouse. In a different study, Walter et al. (2006) reported that the G + C content was similar between imprinted genes and a subset of randomly selected autosomal genes in mouse. The discrepancy between our results and those of Hutter et al. (2006) and Walter et al. (2006) could be due to either species-specific differences or to the small number of imprinted cattle genes in our study.

For tandem repeats and CpG islands, there is accumulating evidence correlating these elements and genomic imprinting. Accordingly, Neumann et al. (1995) suggested using these elements as a search tool for imprinted genes. In this study the average number of CpG islands per kb genomic sequence was significantly higher in the coding sequences of imprinted genes than in control genes. This result was in agreement with the high density of CpG islands reported in imprinted genes in mouse but different than that found for imprinted human genes (Hutter et al. 2006).

For many imprinted genes, monoallelic expression is associated with differentially methylated region (DMR). In a search for sequence elements specific to primary DMRs in the mouse, it has been found that CpG content is higher in DMRs than in whole-genome sequence and in nonimprinted CpG islands (Kobayashi et al. 2006). Also, it has



been found that paternally methylated DMRs have a lower density of CpGs than maternally methylated DMRs (Kobayashi et al. 2006). Of the 14 imprinted cattle genes examined here, 10 of their mouse homologs are associated with DMRs (reviewed in Kobayashi et al. 2006). At present, cattle DMRs are poorly understood. Hence, identification of additional imprinted genes and further investigation of cattle DMRs would improve our understanding of characteristics of imprinted genes.

Similar to CpG islands, we found a notable difference in the abundance of tandem repeats in bovine imprinted genes compared with control genes. This is in agreement with previous reports on the occurrence of these repeats in mouse and human imprinted genes (Hutter et al. 2006; Okamura and Ito 2006; Shirohzu et al. 2004). The high density of CpG islands and tandem repeats observed in the coding sequences of imprinted genes compared with 5' and 3' flanking sequences and to coding and flanking sequences of biallelically expressed genes implies that these elements have an important role in the monoallelic expression of imprinted genes.

Although the mechanisms by which tandem repeats affect genomic imprinting are not currently known, comparative analysis between species could provide a powerful tool to understand these mechanisms. This approach has proven successful in the identification of tandem repeats associated with imprinted genes *Rasgrf1* and *Impact* (reviewed in Okamura and Ito 2006). Of considerable interest was our observation that highly conserved tandem repeats were found in nine imprinted genes in human, mouse, and cattle species. For *GNAS* and *IGF2R*, conserved tandem repeats were found in their CpG islands. Such high conservation indicates that these repeats might have a role in the regulation of allele-specific expression by attracting epigenetic modifications (Hutter et al. 2006; Neumann et al. 1995).

The observation that X-chromosome inactivation is associated with a high concentration of LINES (Lyon 1998) has prompted several research groups to investigate the association of retrotransposable elements with human and mouse imprinted loci. In this study SINE elements were notably fewer in both coding and flanking sequences of imprinted genes compared to control genes. In a search for genomic characteristics that distinguish imprinted from nonimprinted genes in human, Greally (2002) found that the concentration of SINEs was much lower in imprinted loci compared to biallelically expressed genes. In a different study, Allen et al. (2003) found that the frequency of SINE sequences was lower in the flanking regions of monoallelic and imprinted genes. It is assumed that SINEs are either readily removed from imprinted regions or they are unable to transpose to these regions (Greally 2002).

LINES were found to be significantly underrepresented in coding sequences of bovine imprinted genes compared with control genes, in contrast to Walter et al. (2006) who found that LINE elements were significantly denser in imprinted genes with a G + C content of greater than 40% compared with nonimprinted genes in the mouse. On the other hand, analysis of total coding and flanking sequences revealed that LINE-1 frequencies were not statistically different between imprinted and other gene groups examined. In addition, LINE-1 frequency was higher in the combined coding and flanking regions of imprinted genes (10.7%) than in coding sequences alone (4.7%). A high frequency of LINE-1s was observed in flanking regions of human and mouse monoallelically expressed genes but not necessarily in the regions of imprinted genes (Allen et al. 2003). It is conceivable that the distribution of LINES in regions of imprinted genes could be species-specific.

Data obtained in this study showed that about a 8.8-Mb sequence of bovine X chromosome has 26.8% LINES compared to 13.7% found in the autosomal nonimprinted genes. That LINE densities in cattle imprinted genes were low but were high in bovine X chromosome suggests that not necessarily the same mechanisms control X-chromosome inactivation and imprinting in cattle. It has been suggested that LINES have a role in X-chromosome inactivation based on the density of these elements in the X chromosome (Lyon 2006). In fact, it has been found that human and mouse X chromosomes have 26% and 28.5% LINES compared with 13% and 14.6% LINE sequences in autosomal sequences, respectively (Lyon 2006). Thus, the high density of LINES in the bovine X chromosome could confirm the hypothesis that LINES are a common feature of mammalian X chromosomes and that these elements have a function in X-chromosome inactivation (Lyon 2006).

In summary, in this study we investigated the imprinting status of 22 genes in human, mouse, and cattle and found that only 11 genes were conserved across the three species, of which seven genes were paternally expressed and three were maternally expressed. Comparison of sequence characteristics between imprinted and non-imprinted cattle genes revealed that coding sequences of imprinted genes have a higher G + C content and more CpG islands and tandem repeats than biallelically expressed genes. In contrast, imprinted genes have a lower concentration of retrotransposable elements compared with control genes. Of particular interest was the finding of conserved tandem repeat sequences across the three species, which indicates that these elements may have a role in the regulation of imprinting. Taken together, these sequence characteristics could be employed in the prediction of imprinted genes.

**Acknowledgments** The authors thank Jenifer Cruickshank for helpful comments. This study was supported by USDA Hatch grant No. WIS04895 from the University of Wisconsin-Madison.

## References

- Allen E, Horvath S, Tong F, Kraft P, Spiteri E, et al. (2003) High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci U S A* 100:9940–9945
- Dindot SV, Kent KC, Evers B, Loskutoff N, Womack J (2004) Conservation of genomic imprinting at the XIST, IGF2, and GTL2 loci in the bovine. *Mamm Genome* 15:966–974
- Engemann S, Strodicke M, Paulsen M, Franck O, Reinhardt R, et al. (2000) Sequence and functional comparison in the Beckwith-Wiedemann region: implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet* 9:2691–2706
- Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
- Greally JM (2002) Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* 99:327–332
- Higashimoto K, Soejima H, Yatsuki H, Joh K, Uchiyama M, et al. (2002) Characterization and imprinting status of *OBPH1/Obph1* gene: implications for an extended imprinting domain in human and mouse. *Genomics* 80:575–584
- Hutter B, Helms V, Paulsen M (2006) Tandem repeats in the CpG islands of imprinted genes. *Genomics* 88:323–332
- Khatib H (2004) Imprinting of Nesp55 gene in cattle. *Mamm Genome* 15:663–667
- Khatib H (2005a) The COPG2, DCN, and SDHD genes are biallelically expressed in cattle. *Mamm Genome* 16:545–552
- Khatib H (2005b) Characterization and analysis of the imprinting status of the ovine SDHD and COPG2 genes. *Anim Genet* 36:186–188
- Killian JK, Nolan CM, Wylie AA, Li T, Vu TH, et al. (2001) Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. *Hum Mol Genet* 10:1721–1728
- Kim J, Bergmann A, Lucas S, Stone R, Stubbs L (2004) Lineage-specific imprinting and evolution of the zinc-finger gene ZIM2. *Genomics* 84:47–58
- Kobayashi H, Suda C, Abe T, Kohara Y, Ikemura T, et al. (2006) Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs. *Cytogenet Genome Res* 113:130–137
- Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, et al. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat Genet* 36:1291–1295
- Lyon MF (1998) X-chromosome inactivation spreads itself: effects in autosomes. *Am J Hum Genet* 63:17–19
- Lyon MF (2006) Do LINES have a role in X-chromosome inactivation? *J Biomed Biotechnol* 2006(1):1–6
- Mizuno Y, Sotomaru Y, Katsuzawa Y, Kono T, Meguro M, et al. (2002) Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem Biophys Res Commun* 290:1499–1505
- Monk D, Arnaud P, Apostolidou S, Hills FA, Kelsey G, et al. (2006) Limited evolutionary conservation of imprinting in the human placenta. *Proc Natl Acad Sci U S A* 103:6623–6628
- Morison IM, Ramsay JP, Spencer HG (2005) A census of mammalian imprinting. *Trends Genet* 21:457–465
- Neumann B, Kubicka P, Barlow DP (1995) Characteristics of imprinted genes. *Nat Genet* 9:12–13
- Okamura K, Ito T (2006) Lessons from comparative analysis of species-specific imprinted genes. *Cytogenet Genome Res* 113:159–164
- Paulsen M, El-Maarri O, Engemann S, Strodicke M, Franck O, et al. (2000) Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum Mol Genet* 9:1829–1841
- Ruddock NT, Wilson KJ, Cooney MA, Korfiatis NA, Tecirlioglu RT, et al. (2004) Analysis of imprinted messenger RNA expression during bovine preimplantation development. *Biol Reprod* 70:1131–1135
- Shirohzu H, Yokomine T, Sato C, Kato R, Toyoda A, et al. (2004) A 210-kb segment of tandem repeats and retroelements located between imprinted subdomains of mouse distal chromosome 7. *DNA Res* 11:325–334
- Walter J, Hutter B, Khare T, Paulsen M (2006) Repetitive elements in imprinted genes. *Cytogenet Genome Res* 113:109–115
- Weidman JR, Murphy SK, Nolan CM, Dietrich FS, Jirtle RL (2004) Phylogenetic footprint analysis of IGF2 in extant mammals. *Genome Res* 14:1726–1732
- Wood AJ, Roberts RG, Monk D, Moore GE, Schulz R, et al. (2007) A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet* 3:e20
- Zaitoun I, Khatib H (2006) Assessment of genomic imprinting of SLC38A4, NNAT, NAP1L5, and H19 in cattle. *BMC Genet* 7:49
- Zhang S, Kubota C, Yang L, Zhang Y, Page R, et al. (2004) Genomic imprinting of H19 in naturally reproduced and cloned cattle. *Biol Reprod* 71:1540–1544