ORIGINAL ARTICLE

Distribution and characteristics of bovine leukemia virus integration sites in the host genome at three different clinical stages of infection

T. Miyasaka · K. Oguma · H. Sentsui

Abstract Bovine leukemia virus (BLV) is an oncogenic retrovirus closely related to human T-cell lymphotropic virus. BLV-infected cattle are categorized as asymptomatic carriers or as having persistent lymphocytosis or enzootic bovine leukemia, depending on the clinical stage. We investigated the BLV integration site distribution at three BLV clinical stages and examined genome sequence features around the integration sites. In all, 264 BLV integration sites, at various locations on each chromosome, were identified in 28 cattle by inverse PCR and BLAST searches. Approximately one-third of BLV proviruses were independently integrated within transcriptional units, and approximately 10 % were integrated near transcription start sites. Moreover, less than 7 % of BLV integration sites were located near CpG islands. BLV did not preferentially integrate into transcriptionally active regions during any of the clinical stages. At the nucleotide level, regions around BLV integration points were significantly A/T rich with weak sequence consensus. BLV preferentially integrated within long interspersed nuclear repeat elements. Although BLV integration sites may not be associated with disease progression, integration is selective at the nucleotide level.

Introduction

Bovine leukemia virus (BLV) is the etiologic agent of enzootic bovine leukosis (EBL) and belongs to the genus *Deltaretrovirus* (family *Retroviridae*), together with human

T. Miyasaka · K. Oguma · H. Sentsui (✉)
School of Veterinary Medicine, Nihon University,
1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan
e-mail: sentsui.hiroshi@nihon-u.ac.jp

T-cell lymphotropic virus (HTLV). Hence, these two viruses are considered to have similar life cycles and pathogenic strategies. BLV infection has become widespread globally, causing severe economic losses in the livestock industry [25, 27]. However, no effective vaccine or therapy has been developed to date. EBL induction is a pathologically complex process, involving numerous interactions between the host and virus. BLV-induced bovine leukosis is categorized into three different stages [22]. Although most BLV-infected cattle remain asymptomatic (aleukemic; AL), approximately one-third of them exhibit persistent lymphocytosis (PL), which is a polyclonal proliferation of non-neoplastic B lymphocytes. In general, PL is a prominent risk factor for tumor development, but PL cattle are diagnosed as clinically healthy. After latency for 3–5 years, less than 5 % of BLV-infected cattle develop EBL, which is characterized as a neoplastic B cell lymphoma/leukemia.

Integration is one of the most unique features of retroviruses and also a crucial step in their life cycle [19]. After a retrovirus enters a host cell, complementary DNA is synthesized from the viral RNA genome by reverse transcriptase, using transfer RNA as a primer. Subsequently, this DNA is inserted into host genomic DNA by viral integrases and other cellular cofactors. Recent studies have revealed that each retrovirus has distinct integration patterns. For example, human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) show a strong bias for integration within transcription units and transcription start sites [8, 11, 29]. Likewise, murine leukemia viruses (MLVs) favor integration near transcription start sites and CpG islands [31, 33]. In contrast, avian sarcoma/leukosis viruses show no preference for any of these particular locations [2, 26]. It is well known that integration site selection in the host genome is strongly associated with

disease progression. Indeed, HTLV integration sites are more numerous in transcriptionally active regions during pathogenic events such as adult T-cell leukemia (ATL) or HTLV-associated myelopathy/tropical spastic paraparesis (HAM/TSP) [9, 21]. Moreover, retrovirus vector therapy can induce unexpected oncogene activation and development of leukemia in humans [4, 31]. In these cases, MLV vectors are integrated near the start site of the gene for LIM-only protein 2 (*LMO2*), which has been associated with T-cell lymphoma. Thus, retroviral integration site selection changes both host and viral gene expression and influences the condition of the host. However, detailed mechanisms for the selection of retroviral integration sites remain unclear.

Several reports have identified BLV integration sites [7, 13–15, 24]. However, most of these studies have focused on the late stages of infection, especially EBL. Therefore, limited data are available for comparisons of BLV integration sites at the three clinical stages, and it remains unclear whether the BLV integration profile plays a role in disease progression. Accordingly, we investigated the BLV integration site distribution in bovine genomes at all three clinical stages and assessed the characteristics of host genome sequences around the integration sites.

Materials and methods

Sample preparation

A total of 28 female Holstein cattle with natural BLV infections were used in this study. Three AL and four PL animals were maintained on a dairy farm in Japan. To diagnose the BLV clinical stage, complete blood cell counts and leukocyte differentials were determined using standard hematologic techniques and were sorted using a standard leukosis key for cattle [18]. Twenty-one EBL samples were obtained from a meat inspection center in Japan. Genomic DNA was extracted from whole blood of AL and PL cattle, and from lymphomas of EBL cattle. DNA extraction was performed using a Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA). BLV infection was confirmed by PCR [1].

Control integration sites were constructed with reference to an HTLV study [17]. In brief, BLV 5' long terminal repeat (LTR) fragments were prepared from the FLK-BLV cell line [32], which is persistently infected with BLV, by PCR using the primer pair LTRic-F (ATGAAAGATCATGCCGACCT) and LTRic-R, (TGGGTCCAGAGAGTTGTTAGG). Purified PCR products were treated with Klenow DNA polymerase. High-molecular-weight DNA derived from the MDBK cell line was digested with *Hae*III, *Pvu*II, and *Eco*RV and then purified. Digested DNA was ligated with the 5' LTR using T4

DNA ligase (New England Biolabs, Beverly, MA). These artificially generated samples were also subjected to inverse PCR for use as controls.

Inverse PCR

Inverse PCR was performed as described previously with some modifications [24]. In brief, genomic DNA was digested with *Bcl*I or *Mfe*I and was purified. Digested DNA was self-ligated using T4 DNA ligase and used as a template DNA. PCR was conducted using a GoTaq Master Mix (Promega) with the primer pair BMM-F (TACTT TCTGTTTCTCGCGGC) and BMM-R (GACGTCTCTGTC TGGTTTAC). The thermal cycling profile was as follows: initial denaturation for 10 min at 95 °C; 50 cycles of 1 min at 95 °C, 1 min at 61 °C, and 3 min at 72 °C; and final extension for 10 min at 72 °C. PCR products were electrophoresed on agarose gels and were purified from the gels using a Wizard SV Gel and PCR Clean-Up System (Promega Corporation, Madison, WI, USA). Cycle sequence reactions were performed using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Life Technologies, Foster City, CA) with BMM-F or BMM-R primers. Some PCR products were cloned into pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA), and bacterial transformation was performed. These plasmids were used as templates for sequencing.

Mapping of integration sites and statistical analysis

Raw sequences containing the BLV 5' LTR region were selected. The bovine genome sequence adjacent to this 5' LTR was compared using the University of California, Santa Cruz (UCSC) Cow BLAT Search (http://genome. ucsc.edu/cgi-bin/hgBlat?command=start&org=Cow&db= bosTau4&hgsid=164187141) against the November 2009 freeze of the cow genome sequence. All sequences matched the bovine genome with more than 95 % identity. RepeatMasker (http://www.repeatmasker.org/cgibin/WEB RepeatMasker) was used to screen repeat elements.

Statistical analysis was conducted using Fisher's exact and the Mann-Whitney U test. Odds ratios (ORs), 95 % confidence intervals (CIs), and *P*-values were calculated using R version 2.12.2 statistical computing software [30]. Differences were considered significant when *P*-values were less than 0.05.

Results

BLV integration site distribution in host chromosomes

Three AL, four PL, and 21 EBL samples were used to investigate BLV integration site distribution. A total of 264
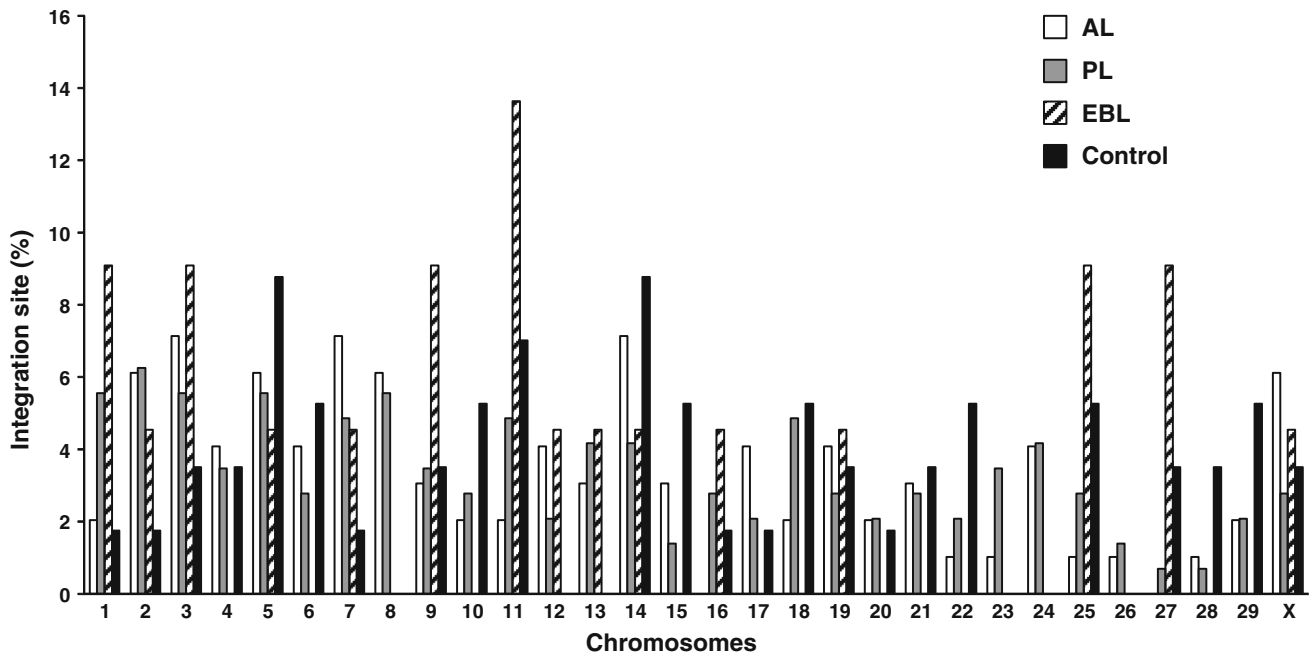
Fig. 1 Chromosomal distribution of BLV integration sites in AL (98 sites), PL (144 sites), EBL (22 sites), and control (57 sites) samples; populations of integration sites in bovine chromosomes are shown as percentages

distinct integration sites, including 98 from AL, 144 from PL, and 22 from EBL samples, that were adjacent to the BLV 5′ LTR were identified in 28 animals. Multiple integration sites were detected in AL and PL cattle, since they possessed a large number of polyclonal infected cells. Some of these integration sites were repeatedly detected in AL and PL cattle. In contrast, EBL individuals harbored only one copy of provirus in tumor cells, except for one sample, which contained two distinct integration sites. For comparison with clinical samples, controls were constructed using *Hae*III, *Pvu*II, and *Eco*RV digestion of DNA derived from MDBK cells and ligation of a 5′ LTR PCR product as reported previously [17]. These were analyzed by inverse PCR, and 57 distinct control integration sites were observed.

Initially, we mapped chromosomal locations of BLV integration sites. Figure 1 shows the distribution of BLV provirus in bovine chromosomes. BLV integration sites were distributed in a variety of areas in every chromosome, with no concentration in specific genome regions. Although one PL sample and one AL sample possessed proviral insertions within the zinc finger protein 394 gene (*ZNF394*) on chromosome 25, integration sites in these samples were at different nucleotide positions.

Preferential location of BLV provirus in the host genome

To investigate the location of BLV integration sites in host genomes, we determined whether BLV integration sites

Table 1 Characteristics of detected integration sites

	BLV integration sites						Control	
	AL (3 animals)		PL (4 animals)		EBL (21 animals)			
	n	%	n	%	n	%	n	%
Total integration sites detected	98	100.0	144	100.0	22	100.0	57	100.0
Transcription unit	27	27.6	48	33.3	8	36.4	18	31.6
Intron	26	26.5	47	32.6	7	31.8	16	28.1
Exon	1	1.0	1	0.7	1	4.5	2	3.5
Sense orientation*	12	44.4	23	47.9	2	25.0	10	55.6

* BLV provirus integrated in the sense orientation within the transcription unit

were located within transcription units (Table 1), i.e., between transcription start and termination sites. The frequencies of BLV integration sites within transcription units were 27.6 % (27 of 98 sites) in AL, 33.3 % (48 of 144 sites) in PL, 36.4 % (8 of 22 sites) in EBL, and 31.6 % (18 of 57 sites) in control samples. Thus, no apparent differences were observed between the three BLV clinical stages and the control. Most BLV integration sites within transcription units were located in introns, while only three integration sites were located in exons.

Fig. 2 Distances from the nearest transcription start sites

Fig. 3 Percentages of sites within 1 kb of CpG islands
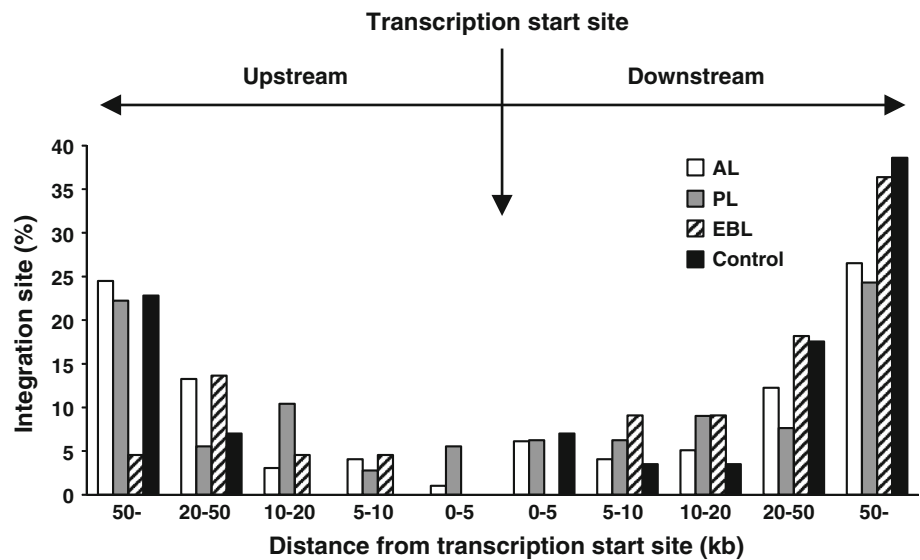
Gene orientations of BLV integration sites within transcriptional units were examined (Table 1). When the transcriptional orientation of BLV provirus was the same as that of the host gene, the BLV integration site was said to be sense oriented. BLV integration sites in the sense orientation were 44.4 % (12 of 27 sites) for AL, 47.9 % (22 of 48 sites) for PL, 25.0 % (2 of 8 sites) for EBL, and 55.6 % (10 of 18 sites) for control samples. Although the difference was not statistically significant, BLV proviruses in the sense orientation were fewer in EBL samples than in the control.

Because the distance from transcription start sites is important for the transcriptional activity of retrovirus genes, we analyzed the distance from the nearest transcription start sites. Figure 2 shows the BLV integration site distribution in BLV clinical stages and control samples. The frequencies of BLV provirus integration within 5 kb of a transcription start site were 7.1 % (7 of 98 sites) in AL, 11.8 % (17 of 144 sites) in PL, and 0 % (0 of 22 sites) in EBL samples. These percentages were not different from those of the control samples (7.0 %, 4 of 57 sites).

CpG islands are rich in C/G dinucleotides and are related to gene regulatory regions such as promoters. Figure 3 summarized the percentages of BLV proviruses that were integrated within 1 kb of CpG islands in three BLV clinical stage and control samples. The ratio of BLV integration sites near CpG islands was 5.1 % (5 of 98 sites) in AL, 4.2 % (6 of 144 sites) in PL, and 4.5 % (1 of 22 sites) in EBL, and no BLV integration sites were located inside CpG islands. In contrast, 7.0 % (4 of 57 sites) of control integration sites were located near CpG islands, and two of these were located inside CpG islands. Frequencies of BLV integration sites were almost identical among the three clinical stages and did not significantly differ from those of the control.
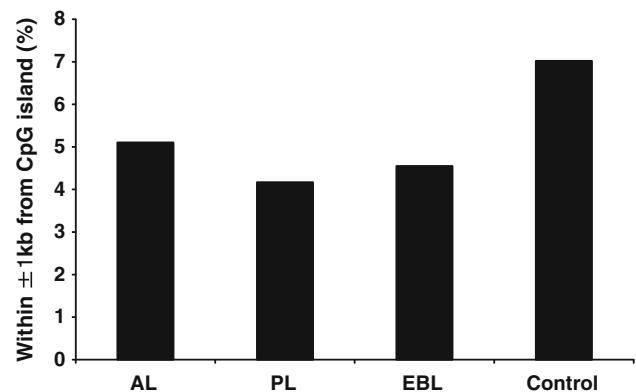
Base arrangements of BLV integration sites

To characterize nucleotide sequences of integration sites, 100-base nucleotide sequences were determined around integration points. A total of 264 BLV integration sites in clinical samples were compared with 57 control integration sites. In clinical samples, average percentages of A, C, G, and T nucleotides were 29.6 %, 20.1 %, 20.7 %, and 29.5 %, respectively, and A and T nucleotides were much more prevalent than C and G. In contrast, the percentages of these nucleotides in control integration sites were 23.8 %, 26.2 %, 26.0 %, and 24.0 %, respectively, indicating fewer A and T nucleotides than C and G nucleotides. Therefore, we confirmed the relative prevalence of the A/T content in BLV integration sites in 96 of 100 positions (96.0 %) (Fig. 4A). The A/T content averaged 59.1 % and ranged from 42.4 to 79.9 % in all BLV integration sites in clinical samples, but averaged 47.8% and ranged from 24.6 to 63.2 % in control integration sites (Fig. 4B). The A/T content of BLV integration sites in clinical samples was

Fig. 4 Distribution of the A/T content in each 100 nucleotides around BLV integration points. (A) A/T percentage of each nucleotide position in 264 BLV and 57 control integration sites. Position 0 indicates the BLV integration point. (B) The average A/T content in 100 nucleotides around BLV integration points from 264 BLV and 57 control integration sites
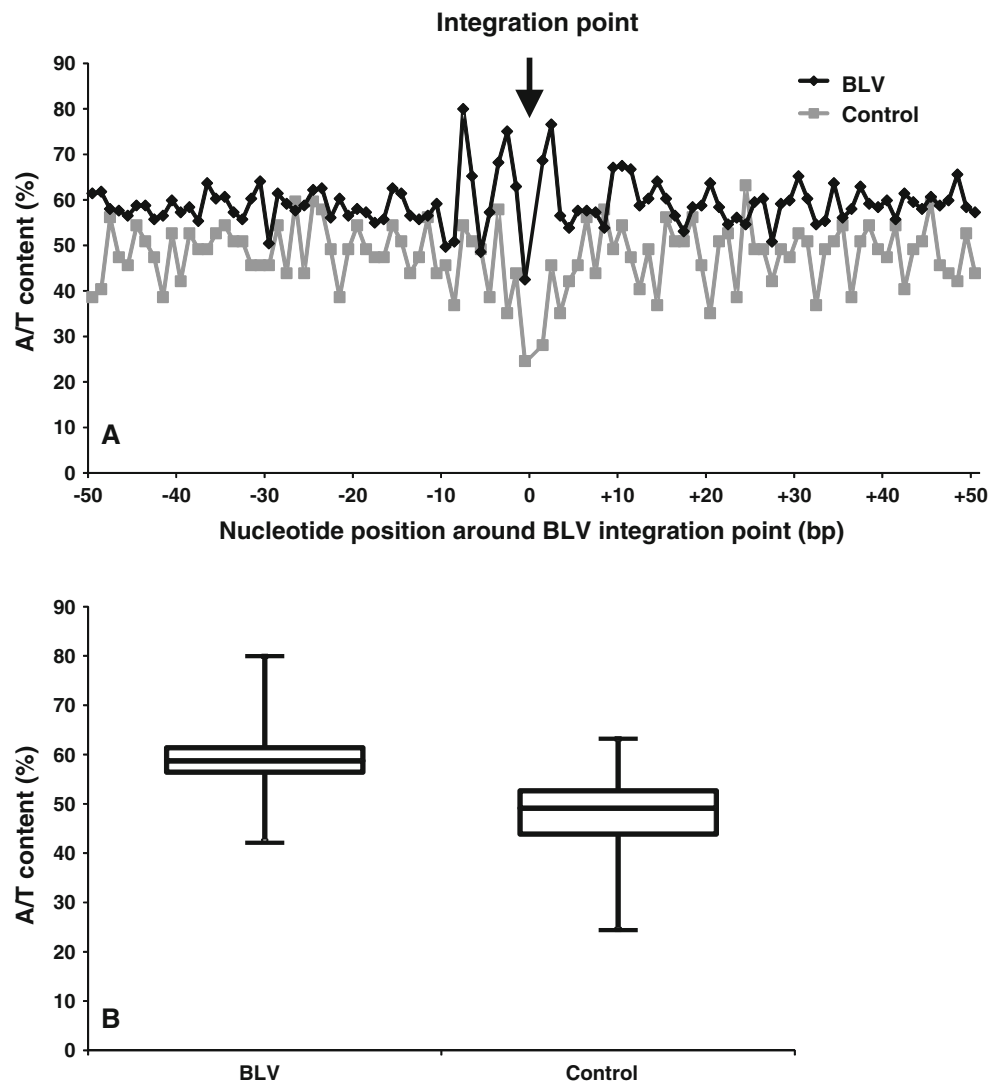
significantly higher than that of control integration sites ($P < 0.0001$). Since these control integration sites were artificially constructed by digestion with restriction enzymes, each of the six bases around the integration point contained restriction enzyme recognition sequences. When these sequences were excluded from the analysis, the A/T content of controls averaged 48.6 % and was not significantly different from that of modified control samples ($P = 0.5511$).

Among BLV integration sites, nucleotide positions at −8, −4, −3, +1, and +2 showed prominent A/T contents of 79.9 %, 68.2 %, 75.0 %, 68.6 %, and 76.5 %, respectively (Fig. 4A). These nucleotide positions corresponded to an excess of T at −8 (46.6 %), A at −4 (37.7 %), A at −3 (44.7 %), T at +1 (40.9 %), and A at +2 (52.0 %; Fig. 5). In contrast, nucleotides at position −1 had a lower A/T content (42.4 %) and an excess of C (34.9 %).

BLV integration into interspersed repeat elements of the host genome

To investigate the association between interspersed repetitive sequences and BLV integration sites, a total of 264 BLV integration sites in clinical samples were compared with 57 control integration sites (Table 2). Integration into repetitive elements occurred in 45.1 % (119 of 264 sites) of clinical samples and 38.6 % (22 of 57 sites) of control samples. BLV integration sites were most frequently observed in long interspersed nuclear elements (LINEs). Integration of BLV into LINEs occurred in 25.8 % (68 of 264 sites) of clinical samples and 12.3 % (7 of 57 sites) of control samples. Hence, BLV integration sites occurred with significantly higher frequency in LINEs in lymphocytes than in controls ($P = 0.0372$). The second most preferential targets for BLV integration sites were short interspersed nuclear elements (SINEs). Integration sites of

Fig. 5 Nucleotide components at each of 50 nucleotides around 264 BLV integration sites
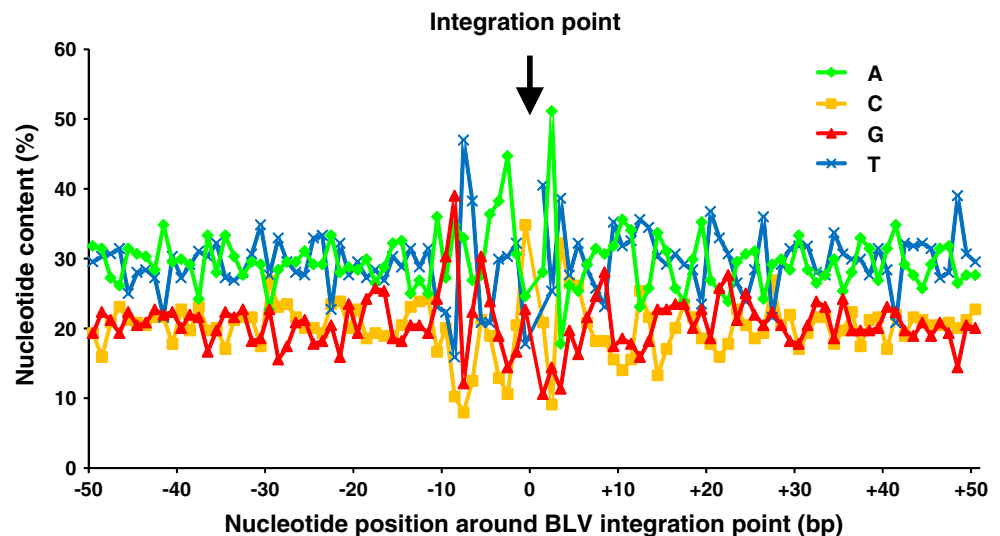

Table 2 BLV integration sites located in repeat elements

	BLV integration sites		Control		OR (95 % CI)	P-value
	n	%	n	%		
Total of integration sites detected	264	100.0	57	100.0		
LINEs	68	25.8	7	12.3	2.5 (1.0–6.8)	0.0372
SINEs	29	11.0	2	3.5	3.4 (0.8–30.1)	0.1337
LTR elements	11	4.2	1	1.8	2.4 (0.3–106.6)	0.7000
DNA elements	3	1.1	2	3.5	0.3 (0.0–3.9)	0.2166
Simple repeats	4	1.5	0	0.0		1
Low complexity	3	1.1	0	0.0		1

BLV in SINEs occurred in 10.5 % (29 of 264 sites) of clinical samples and 3.5 % (2 of 57 sites) of control samples, though no significant difference was observed ($P = 0.1337$). The frequency of BLV integration sites within other elements was less than 5 %, and no significant differences were observed between lymphocytes and controls.

Discussion

Recent genome-wide surveys have demonstrated that retroviral integration selection is not a totally random event. Indeed, retroviruses have distinct integration targets in host genomes [21, 23, 29, 31]. However, comparisons of BLV integration sites at three clinical stages have not been reported to date. In the present study, we investigated the BLV integration site distribution in bovine genomes at AL, PL, and EBL stages. Surprisingly, BLV proviruses were distributed across every chromosome, and no integration hot spot was identified, indicating that unlike MLVs, BLV integration into specific genome regions does not directly cause disease onset [4, 31]. Thus, we investigated the relationship between BLV integration sites and gene regulatory regions. Generally, the transcriptional activity of the virus is closely related to the host genome region, which also influences disease progression. For example, in HTLV, integration sites of HAM/TSP patients were preferentially located in transcription units when compared with those in asymptomatic carriers [21]. HTLV integration sites in ATL patients were also found at much higher rates in transcription start regions than in asymptomatic carriers [9]. Interestingly, although approximately one-third of BLV proviruses were located within transcriptional units, no relationship to gene orientation was found. Moreover, only approximately 10 % of BLV proviruses were found near transcriptional start sites. Less than 7 % of BLV integration sites were located near the CpG islands. These percentages were similar to those in control samples. Taken together, BLV has no preference for integration into transcriptionally active regions. Furthermore, these trends of BLV integration sites in host genomes were common to all three BLV clinical stages. Hence, it is unlikely that BLV integration site selection is associated with disease progression. In general, virus-infected cells with high transcriptional activity tend to be a target for host immune attack [12]. Hence, it is thought that BLV-infected cells with provirus in transcriptionally active regions are eliminated prior to latency. Despite the lack of apparent oncogenes, the BLV genome contains a variety of regulatory proteins, such as Tax and Rex of *pX*, which control the transcriptional activity of both host cells and virus regions [10]. Similarly, U5 regions of LTRs are also responsible

for signal transduction in host cells via Tax-independent mechanisms [16]. These candidate regulatory factors control the transcriptional activity of not only viruses but also host cells, and contribute to the maintenance of viral load. Hence, there may be no need for BLV integration into transcriptionally active regions because the virus has its own regulatory systems.

In subsequent experiments, we determined the nucleotide composition of BLV integration sites and found that the integration sites were rich in certain nucleotides. Retrovirus integration sites are mainly determined by viral integrases, which choose the integration site by interacting with chromatin structure, target DNA sequences, and other viral and cellular factors [3]. For example, a recent study revealed that lens epithelium-derived growth factor (LEDGF/p75) interacted with both human genomes and HIV integrases and affected integration site selection [6, 20]. In their studies, knockdown of LEDGF/p75 dramatically modulated target selection for HIV not only at the gene level but also at the nucleotide level. Thus, as a result of LEDGF/p75 depletion, HIV integration was less frequent in transcriptional units and LEDGF/p75-associated genes, and was more frequent in A/T-rich sequences. In our analyses, BLV had a significant base preference for integration into A/T-rich sequences. Moreover, a number of nucleotide positions, such as -8, -4, -3, -1, $+1$, and $+2$, had weakly conserved sequences. Accordingly, it is conceivable that these nucleotide positions are recognized by the BLV integrase and its cofactors that mediate BLV integration. HTLV integration is also known to have a strong bias for A/T-rich sequences and slight consensus sequence motifs have been observed [5, 17]. However, information on BLV and HTLV integrases and their interacting host factors is limited. A previous study suggested that the BLV genome structure is similar to that of HTLV and that over 70 % of their genomes are related at the nucleotide level [28]. In particular, the *pol* regions that encode virus integrases showed close similarity between BLV and HTLV. Therefore, it is speculated that these two viruses use similar mechanisms for integration into host genomes.

The current study revealed the BLV integration site distribution at three BLV clinical stages and examined genome sequence features around integration sites. BLV integration sites were uncharacteristic throughout the course of disease. However, BLV has preferential integration sites at the nucleotide level.

References

1. Asfaw Y, Tsuduku S, Konishi M, Murakami K, Tsuboi T, Wu D, Sentsui H (2005) Distribution and superinfection of bovine leukemia virus genotypes in Japan. Arch Virol 150:493–505
2. Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. J Virol 79:12035–12044
3. Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, Hoffmann C (2005) Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol 3:848–858
4. Check E (2002) A tragic setback. Nature 420:116–118
5. Chou KS, Okayama A, Su IJ, Lee TH, Essex M (1996) Preferred nucleotide sequence at the integration target site of human T-cell leukemia virus type I from patients with adult T-cell leukemia. Int J Cancer 65:20–24
6. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11:1287–1289
7. Couez D, Deschamps J, Kettmann R, Stephens RM, Gilden RV, Burny A (1984) Nucleotide sequence analysis of the long terminal repeat of integrated bovine leukemia provirus DNA and of adjacent viral and host sequences. J Virol 49:615–620
8. Crise B, Li Y, Yuan C, Morcock DR, Whitby D, Munroe DJ, Arthur LO, Wu X (2005) Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. J Virol 79:12199–12204
9. Doi K, Wu X, Taniguchi Y, Yasunaga J, Satou Y, Okayama A, Nosaka K, Matsuoka M (2005) Preferential selection of human T-cell leukemia virus type I provirus integration sites in leukemic versus carrier states. Blood 106:1048–1053
10. Gillet N, Florins A, Boxus M, Burteau C, Nigro A, Vandermeers F, Balon H, Bouzar AB, Defoiche J, Burny A, Reichert M, Kettmann R, Willems L (2007) Mechanisms of leukemogenesis induced by bovine leukemia virus: prospects for novel anti-retroviral therapies in human. Retrovirology 4:18
11. Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, Pierson TC, Margolick JB, Siliciano RF, Siliciano JD (2004) Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J Virol 78:6122–6133
12. Kattan T, MacNamara A, Rowan AG, Nose H, Mosley AJ, Tanaka Y, Taylor GP, Asquith B, Bangham CR (2009) The avidity and lytic efficiency of the CTL response to HTLV-1. J Immunol 182:5723–5729
13. Kettmann R, Meunier-Rotival M, Cortadas J, Cuny G, Ghysdael J, Mammerickx M, Burny A, Bernardi G (1979) Integration of bovine leukemia virus DNA in the bovine genome. Proc Natl Acad Sci USA 76:4822–4826
14. Kettmann R, Cleuter Y, Mammerickx M, Meunier-Rotival M, Bernardi G, Burny A, Chantrenne H (1980) Genomic integration of bovine leukemia provirus: comparison of persistent lymphocytosis with lymph node tumor form of enzootic. Proc Natl Acad Sci USA 77:2577–2581
15. Kettmann R, Deschamps J, Couez D, Claustriaux JJ, Palm R, Burny A (1983) Chromosome integration domain for bovine leukemia provirus in tumors. J Virol 47:146–150
16. Kiermer V, Van Lint C, Briclet D, Vanhulle C, Kettmann R, Verdin E, Burny A, Droogmans L (1998) An interferon regulatory factor binding site in the U5 region of the bovine leukemia virus long terminal repeat stimulates Tax-independent gene expression. J Virol 72:5526–5534
17. Leclercq I, Mortreux F, Cavrois M, Leroy A, Gessain A, Wain-Hobson S, Wattel E (2000) Host sequences flanking the human

T-cell leukemia virus type 1 provirus in vivo. J Virol 74:2305–2312

18. Levy D, Deshayes L, Guillemain B, Parodi AL (1977) Bovine leukemia virus specific antibodies among French cattle. I. Comparison of complement fixation and hematological tests. Int J Cancer 19:822–827

19. Lewinski MK, Bushman FD (2005) Retroviral DNA integration–mechanism and consequences. Adv Genet 55:147–181

20. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM (2006) An essential role for LEDGF/p75 in HIV integration. Science 314:461–464

21. Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CR (2008) HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. PLoS Pathog 4:e1000027

22. Mirsky ML, Olmstead CA, Da Y, Lewin HA (1996) The prevalence of proviral bovine leukemia virus in peripheral blood mononuclear cells at two subclinical stages of infection. J Virol 70:2178–2183

23. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2:E234

24. Murakami H, Yamada T, Suzuki M, Nakahara Y, Suzuki K, Sentsui H (2011) Bovine leukemia virus integration site selection in cattle that develop leukemia. Virus Res 156:107–112

25. Murakami K, Kobayashi S, Konishi M, Kameyama K, Yamamoto T, Tsutsui T (2011) The recent prevalence of bovine leukemia virus (BLV) infection among Japanese cattle. Vet Microbiol 148:84–88

26. Narezkina A, Taganov KD, Litwin S, Stoyanova R, Hayashi J, Seeger C, Skalka AM, Katz RA (2004) Genome-wide analyses of avian sarcoma virus integration sites. J Virol 78:11656–11663

27. Rodriguez SM, Florins A, Gillet N, de Brogniez A, Sanchez-Alcaraz MT, Boxus M, Boulanger F, Gutierrez G, Trono K, Alvarez I, Vagnoni L, Willems L (2011) Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. Viruses 3:1210–1248

28. Sagata N, Yasunaga T, Ohishi K, Tsuzuku-Kawamura J, Onuma M, Ikawa Y (1984) Comparison of the entire genomes of bovine leukemia virus and human T-cell leukemia virus and characterization of their unidentified open reading frames. EMBO J 3:3231–3237

29. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110:521–529

30. Team RDC (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

31. Tsukahara T, Agawa H, Matsumoto S, Matsuda M, Ueno S, Yamashita Y, Yamada K, Tanaka N, Kojima K, Takeshita T (2006) Murine leukemia virus vector integration favors promoter regions and regional hot spots in a human T-cell line. Biochem Biophys Res Commun 345:1099–1107

32. Van der Maaten MJ, Miller JM (1976) Reprication of bovine leukemia virus in monolayer cell cultures. Bibl Haematol 43:360–362

33. Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300:1749–1751