

Target Site Preferences of Subgroup C Rous Sarcoma Virus Integration into the Chicken DNA

Markéta Reinišová¹, Adam Pavlíček^{1,§}, Petr Divina^{1,2}, Josef Geryk¹, Jiří Plachý¹ and Jiří Hejnar^{*,1}

¹*Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videňská 1083, CZ-14220, Prague 4, Czech Republic*

²*Center for Applied Genomics, Videňská 1083, CZ-14220, Prague 4, Czech Republic*

[§]*Present address: Adam Pavlíček, Computational Biology, Pfizer Global R&D, San Diego, CA, USA*

Abstract: We sequenced unselected integration sites of subgroup C Rous sarcoma virus from infected chicken cells and mapped them on the chicken draft genome assembly. Our genome-wide analysis demonstrates the non-random character of subgroup C Rous sarcoma virus integration into genes, gene-rich regions and GC-rich regions. Within genes, there is no significant integration bias in favor of transcription start sites. Integration sites are underrepresented on microchromosomes. These results may be important for the development of gene transfer vectors and gene therapy strategies based on avian sarcoma and leukosis viruses.

INTRODUCTION

Retroviruses replicate through a provirus, the DNA intermediate of their RNA genome integrated into the host genome. Provirus integration is mediated by a virus-encoded enzyme, integrase, which enters the infected cell as a part of the preintegration complex (PIC) and catalyzes the recombination between linear proviral DNA and host DNA. Retrovirus integration potentially occurs in many genomic regions (reviewed in [1]), but not completely randomly. Different genomic regions can be targeted at strikingly different frequencies depending on the local secondary DNA structure, bending, distortion and nucleosome distribution (reviewed in [2]). The mechanism of target site selection remains to be elucidated; to date, the tethering of the PIC with chromosomal DNA mediated by proteins binding both the retroviral integrase and host DNA seems to be the most probable explanation for site-preference of both retrovirus and retrotransposon integration (reviewed in [3]). For example, the proteins of the INI/SNF5 complex and lens epithelium-derived growth factor (LEDGF/p75) have been shown to interact with human immunodeficiency virus type 1 (HIV-1) integrase [4,5]. Depletion of LEDGF/p75 significantly alters the distribution of HIV-1 integration sites [6].

The availability of a representative and unbiased set of HIV-1 integration sites [7] together with the knowledge of the complete human genome sequence enabled us to study the genome-wide distribution of HIV-1 proviruses with respect to the genomic features of the host DNA such as GC content, gene density, and cytogenetic bands. We have shown that HIV-1 preferentially integrates into genes, GC-rich regions and cytogenetic R bands [8]. In a more

extensive study, Schroeder *et al.* [9] added that transcriptionally active genes are favored as integration targets. Using naked genomic DNA and isolated PICs from infected cells in *in vitro* integration assays, the integration bias for GC-rich regions [10] but not for active genes [9] was shown. This finding points to the importance of native chromatin in the interaction with the PIC.

In comparison with HIV-1, murine leukemia virus (MLV) displays only weak preference for actively transcribed genes but favors integration in promoter sequences or in close proximity to transcription start sites [11]. Distinct target site preferences of these retroviruses are retained by HIV- and MLV-derived vectors [12,13]. These results may have relevance for the design of retroviral vectors for gene therapy: in a cohort of child patients treated for X-linked severe combined immunodeficiency with a MLV-based vector, two cases of B cell lymphoproliferative disease occurred when the vector integrated close to the promoter of the LMO2 protooncogene [14]. The clonal selection in favor of retroviruses integrated into or near growth-related genes probably took place during the multistep carcinogenesis. Similarly, in cells from AIDS patients, HIV-1 proviruses were localized almost exclusively in introns of transcribed genes associated with apoptosis, signal transduction and transcriptional control [15] and, predominantly, in GC-rich chromosomal segments [16]. The resting CD4⁺ T cell population from HIV-1-infected individuals also contains more than 90% of proviruses integrated into introns of transcription units, randomly with respect to their transcriptional orientation [17]. Cloning of human T cell leukemia virus type 1 (HTLV-1) integration sites from patients with adult T cell leukemia displayed preference for genes related to cell growth [18,19].

Differences in target site selection by HIV-1 and MLV attract attention to other retroviruses, particularly avian sarcoma and leukosis viruses (ASLV). *In vitro* integration as-

*Address correspondence to this author at the Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videnska 1083, CZ-14220, Prague 4 Academy of Sciences of the Czech Republic, Czech Republic; E-mail: hejnar@img.cas.cz

says showed that ASLV integrase increases its activity on the model target chromatin compacted by addition of histone 1. In contrast, HIV-1 integrase activity is suppressed by nucleosome compaction [20]. Together with previous data on the Rous sarcoma virus (RSV) propensity to epigenetic silencing in mammalian cells [21-23] it can be assumed that ASLV target site selection differs from that of HIV-1 and MLV. In the human genome, ASLV integrates into transcription units with similarly weak preference to MLV, but without any bias for transcription start sites [13,24]. Transcriptionally active genes are slightly favored by ASLV, but no preference for highly expressed genes was found [24].

The basic compositional properties of avian genomes are comparable to mammalian genomes [25-27]. In both the human and chicken genome, gene-rich regions are concentrated in GC-rich isochores. Compositional mapping of chicken mitotic chromosomes revealed that the GC-richest regions are preferentially localized on a large number of microchromosomes and also on almost all telomeric bands of macrochromosomes, whereas the GC-poorest isochores are generally localized on the internal regions of macrochromosomes and are almost absent from microchromosomes [28]. The highest gene and CpG island density, GC and CpG contents are found on microchromosomes shorter by two orders of magnitude than macrochromosomes [29]. On the other hand, the chicken genome is threefold smaller than the human genome at the expense of intergenic regions, intronic sequences, and interspersed repeat content. These important differences between genomes might exert distinct patterns of provirus distribution. Barr *et al.* [30] described the genome-wide integration pattern of subgroup A-ASLV-derived replication-competent vector and HIV-1 in the chicken genome. They demonstrated that ASLV vector weakly and HIV-1 strongly favored integration in active transcription units but without any striking preference for transcription start regions. The complex of ASLVs comprises several subgroups with different receptor specificities of their envelopes and with significant differences among *gag* and *pol* genes. Particularly, the integrase drives the target site preference by tethering with chromatin components [3] and small changes of its structure could induce substantial changes in retrovirus integration. In order to get further data on ASLV integration in its natural host, we have investigated the subgroup C RSV integration specificity in chicken cells and present here a genome-wide analysis of 57 integration sites with respect to the chromosomal distribution, GC content, and transcription features of targeted regions.

MATERIALS AND METHODOLOGY

Cell Infections

Chicken embryo fibroblasts (CEF) were prepared by standard procedures from 10-day-old chicken embryos of the outbred line P, which is free of ASLV-related endogenous sequences [31]. The chicken line P was bred at the Institute of Molecular Genetics, Prague, Czech Republic. CEFs were grown in F-12 and MEM-D mixed 1:1 (Sigma) supplemented with 1% NaHCO₃, 5% calf serum, 5% fetal calf serum, 1% chicken serum, penicillin-streptomycin mix (100 µg/ml each), and 2.5 mg/ml amphotericin B at 39°C and 5% CO₂. High-titer infectious RSV was generated by transfection of pAPrC plasmid DNA containing a complete nonper-

mutated molecular clone of RSV Prague C strain [32]. Transfection was performed using lipofection agent DOTAP (Qiagen) according to the manufacturer's instructions and the viral spread was monitored by observing the cell transformation. Virus stocks were generated by harvesting the cell supernatants, centrifugation at 3,000 x g for 10 min at 4°C, filtration through 0.2 µm cellulose acetate syringe filter (Whatman) and stored at -80°C. Freshly prepared CEFs were infected at 75% confluency with RSV at multiplicity of five by adding the virus stock in a small volume of culture medium to the cells for 1 h. Twenty-four hours after infection, the CEFs were harvested and used for DNA isolation.

Cloning and Sequencing of the Integration Sites

For amplification and cloning of the sequences flanking the 5'LTR of RSV proviruses, we used the I-PCR strategy schematically shown in Fig. (1). Briefly, 1.5 µg of DNA from infected CEFs was digested overnight with *Bst*YI (New England Biolabs) at 60°C and circularized by self-ligation with 6 units of T4 DNA ligase (Roche) in 100 µl reaction overnight at increasing temperature, which started at 10°C and finished at 18°C. The product of self-ligation was subsequently cleaved overnight at 37°C with 8 units of *Eco*81I (Fermentas) in 50 µl reaction and 5 units of *Pvu*II (New England Biolabs) in 40 µl reaction to eliminate the background from internal proviral *Bst*YI fragments and to increase the efficiency of I-PCR by linearization of the circles within LTR, respectively. After desalting, we subverted 150 ng of the resulting DNA to PCR with Taq polymerase (TaKaRa) in standard reaction according to manufacturer's instructions with addition of betaine and dimethylsulfoxide. The primers used were as follows: forward leader, 5'CCTCATCCGTC TCGCTTATTCG3' (nucleotides 63 to 85 3' to the end of 5'LTR), and backward LTR, 5'CCTTACTACCACCAAT CGCA3' (nucleotides 95 to 115 of LTR). The conditions for I-PCR amplification were 95°C for 3 min, followed by 34 cycles each consisting of 94°C for 20 s, 58°C for 30 s, and 72°C for 150 s, and finally 3 min at 72°C. The I-PCR products corresponding to the size from 0.5 to 2.0 kbp were extracted from the agarose gel using the Qiaex gel extraction kit (Qiagen) and 15% aliquots were ligated into the pGEM-T Easy vector (Promega). Ligation products were transformed into XL-1 Blue MRF' bacteria and resulting colonies were screened by black-white selection using S-gal and by PCR using forward leader and backward LTR primer for the presence of inserts longer than 500 bp. Selected clones were subverted to DNA sequencing using the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

Integration Site Mapping and Gene Identification

Integration sites were mapped onto the chicken draft genome assembly using BLAT [33]. The threshold for localization was ≥99% identity over ≥90% of the length of the integration site. Hits to multiple locations in the genome and to unassembled contigs were discarded. Genes (transcriptional units, TU) targeted by integration as well as genes located within 100 kb of the integration sites were identified from two sources: alignment of chicken mRNA sequences from GenBank to the chicken genome available in the all_mrna table from the UCSC Genome Browser Database [34], and alignment of UniGene representative sequences (Gga.seq.uniq file) to the chicken genome using BLAT. The

gene was considered as targeted by integration if the integration site was located inside a region of the aligned mRNA or UniGene sequence. Similarly, genes were considered to be in the 100 kb vicinity of integration if the corresponding mRNA or UniGene sequences were located in the region 50 kb downstream or 50 kb upstream of the integration site. The 100 kb window was used as an optimal unit for description of genomic features and isochoric composition of chicken chromosomes [28]. UniGene cluster ID was used as the common gene identifier to keep a non-redundant list of genes obtained from the alignment of mRNA and UniGene sequences.

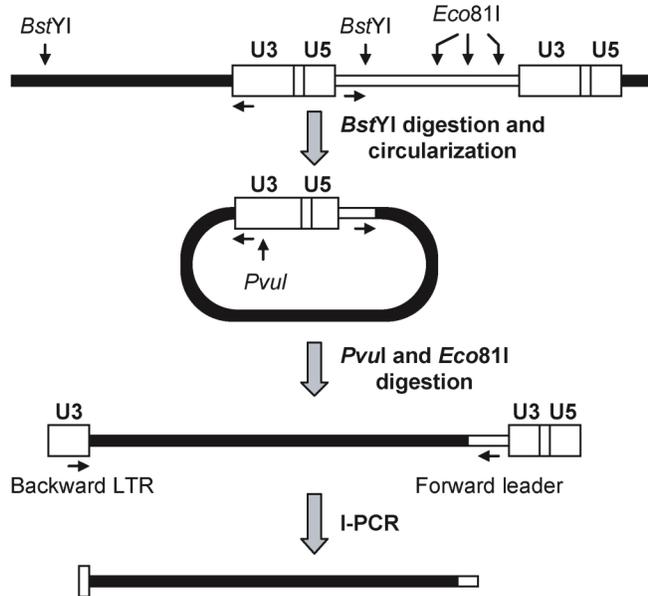


Fig. (1). Scheme for I-PCR to clone the sequences flanking the 5' LTR of RSV integrated into chicken genomic DNA. *BstYI* digest of DNA from infected cells was circularized by self ligation, linearized by *PvuI* and subverted to I-PCR. Enzymes recognition sites are depicted as black-headed vertical arrows, I-PCR primers backward LTR and forward leader as horizontal arrows. Proviral DNA is represented by open boxes, flanking chicken DNA is shown as a filled box. Not exactly to scale.

EST Data Processing

Publicly available chicken EST data were obtained from the NCBI dbEST database [35]. Sixty-three EST libraries including non-normalized and normalized, prepared from bulk tissues with at least 500 ESTs, were selected (Additional data file 2). EST libraries were organized into 14 groups by tissue according to the annotation in the chicken UniGene library browser (<http://www.ncbi.nlm.nih.gov/UniGene/>); and EST data from the same tissue were pooled. For each gene (TU) from the chicken UniGene database the number of ESTs in each of the 14 tissues was enumerated.

Database Versions

The following databases were used in all analyses: chicken draft genome assembly - galGal2, February 2004, UCSC [29] and chicken UniGene - build #27, May 2005 [36].

RESULTS AND DISCUSSION

Cloning and Confirmation of Integration Sites

From two independent infections and inverse PCRs (I-PCR), we obtained in total 91 sequences flanking the RSV 3' LTR. We have taken into account only sequences with the correct structure, i.e. part of the U3 region followed with an unknown genomic, i.e. part of the leader region in proper orientation and with the *BstYI* site on the border between proviral and chicken DNA (Fig. 1). The length of cloned flanking sequences without proviral junctions varied from 36 to 716 nucleotides. Each of the selected 91 sequences was examined by BLAT and the hit that matched the query with highest identity, usually more than 99%, along its full length, was regarded as the integration site. If there were several equally identical hits, the query sequence was regarded as a genomic repeat.

Chromosomal Distribution of Integration Sites

RSV integration sites were localized on chicken chromosomes by their mapping to the assembled genome sequence. In total, 57 out of 91 integration sites were mapped unambiguously to the chicken draft genome assembly (Fig. 2). The rest of integrations landed either in unassembled chromosomes (6 integrations), multiple repeats (15 integrations), or could not be localized in the current version of the genome (13 integrations). We have found proviruses integrated into the majority of macrochromosomes, with the exception of chromosomes 8, 10, 13, 14, 15, 17, and 19. Microchromosomes were mostly untargeted. Chromosomes 1, 2, and 3 hosted 14, 15, and 7 integrations, respectively. The shorter chromosomes 4, 5, 6, 7, 9, 11, 12, 18, and 20 were targeted by 3, 2, 2, 1, 1, 1, 4, 1, and 2 integrations, respectively. There were three hits on chromosome Z and one hit on chromosome W. We have not found any integration hotspot as observed for HIV-1 [9]. RSV integration also shows a significant preference for GC-rich regions (Fig. 3A).

RSV Preferably Integrates into Genes

We next assessed the frequency of RSV integration into genes and intergenic regions. Integration into genes was defined as integrations within annotated chicken mRNA genes (all_mrna) and/or Unigene clusters. Out of 57 localized integrations, 21 (36.8%) were found in TUs (Fig. 2). On the other hand, the analyzed annotated mRNAs and Unigene clusters represent 22.7% of the sequenced chicken genome. This discrepancy is significant ($p = 0.017$; Binomial test) and shows preferential integration of RSV into genes vs intergenic regions. In addition to 21 individual genes targeted by integrations, there is a strikingly increased density of genes around the integration sites (Fig. 3B). This indicates that RSV prefers not only genes, but also gene-rich regions. All identified genes found within 100 kb of the integration sites are listed in Supplementary data file 1.

We were also interested whether RSV inserts randomly into TUs, or displays a preference for some specific regions along the targeted genes. We have combined the Unigene clusters with annotated mRNAs and removed all redundant

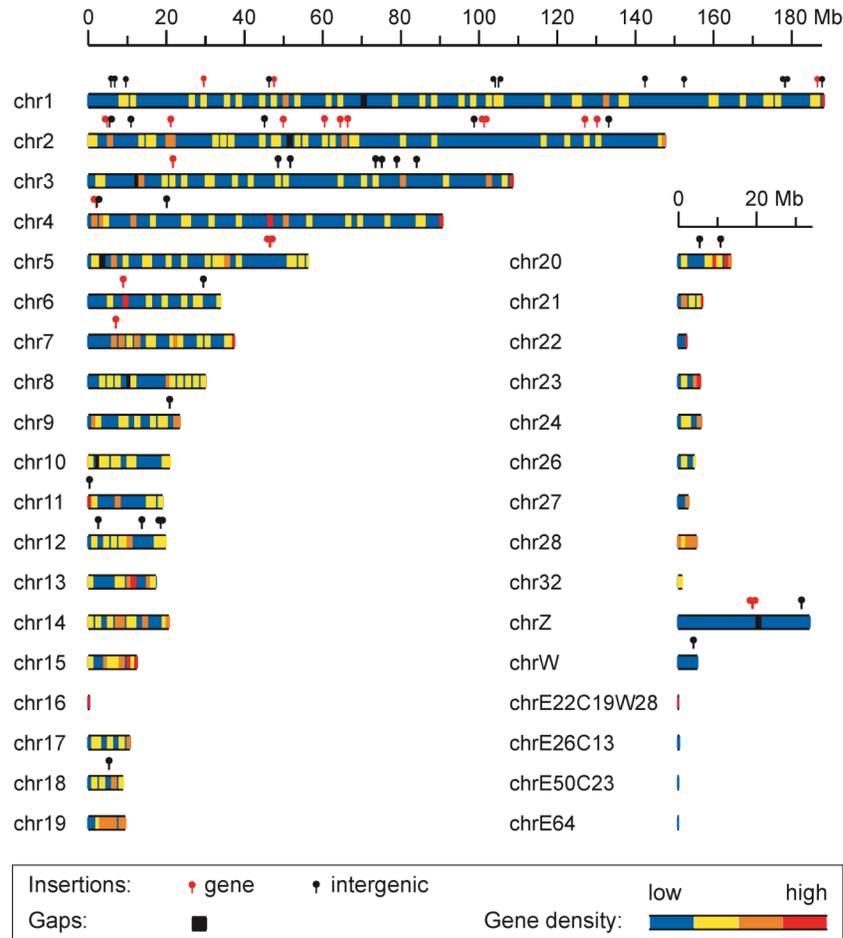


Fig. (2). Chromosomal location of RSV integrations in chicken chromosomes. In total, 57 out of 91 integration sites were unequivocally localized in the assembled chicken genome. Integration into genes, defined as integrations within annotated chicken mRNA genes (all_mrna) and/or Unigene cluster are highlighted as red “lollipops“. Integrations outside of genes are shown in black. Chromosomes are depicted in colors to show gene density over 1 Mb long chromosomal segments, the black regions correspond to long gaps (often putative centromeres). Microchromosomes currently unassembled are not shown.

TUs. The insertions are clustered within 10 kb, but without any further accumulation within 5 kb around the expected transcription start (Fig. 4A). Similarly, analysis of the relative distance from the transcription start shows an increased number of integrations within the first 10% of the gene length but this difference is not significant due to the low number of integrations inspected (Fig. 4B). Taken together, we can conclude only very slight bias for RSV integrations near the transcription start.

RSV Integration Favors Genes Expressed in Multiple Tissues

We investigated the expression of 21 genes targeted by integration and also 164 genes located in the 100 kb vicinity (50 kb downstream and 50 kb upstream) of all integration sites using the publicly available EST data from the NCBI dbEST database. We selected 63 chicken EST libraries corresponding to 14 different tissues (Supplementary data file 2) and classified genes by their expression into three categories: (1) genes with high expression breadth (probably housekeeping genes), which were expressed in 5 or more tissues, (2) tissue-specific genes that were expressed exclusively in one tissue, and (3) genes expressed preferentially in 1 or 2 tissues, which also include the subset of tissue-specific genes.

A gene was considered to be expressed in a tissue if at least one corresponding EST was present in that tissue. The results from the expression analysis are summarized in (Table 1). We observed that the genes expressed in 5 or more tissues represent 42.9% of genes targeted by integration and 39.6% of genes within 100 kb of the integration site (Fig. 5). In comparison to the rest of the genes in UniGene, of which only 27.7 % were expressed in 5 or more tissues, this repre-

sents a significant excess in the 100 kb vicinity of the integration site ($p = 0.001$, Fisher’s exact test). Genes with high expression breadth were also more abundant among the genes targeted by integration, but the difference was not significant due to the low sample size. For the tissue-specific genes, the opposite trend was observed. Only 14.3% and 18.3% of tissue-specific genes were detected among the genes targeted by integration and genes located in the 100 kb vicinity of the integration site, respectively, while 25.7% of the rest of genes in UniGene were tissue-specific. The paucity of tissue-specific genes within 100 kb of the integration site was significant ($p = 0.03$, Fisher’s exact test). Similar results as for the tissue-specific genes were obtained for the genes expressed preferentially in 1 or 2 tissues (Table 1),

which displayed a significantly low frequency in the 100 kb vicinity of the integration site ($p = 0.01$, Fisher's exact test). As far as we can transfer the data from EST libraries to CEFs, whose transcriptome was not yet analysed, the integrations are favored in regions containing genes with high expression breadth (probably housekeeping) and disfavored in the vicinity of genes expressed specifically in one tissue or preferentially in a small number of tissues.

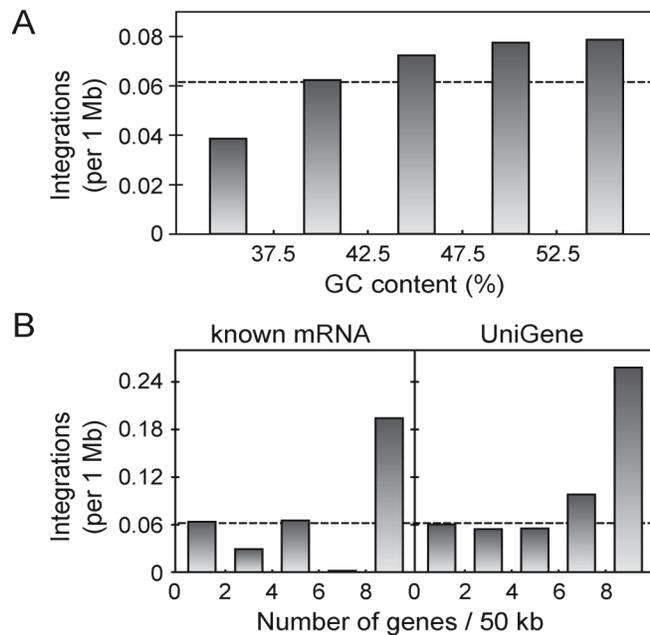


Fig. (3). GC and gene content of integration sites. (A) GC content. The plot shows density of integration according to the GC content in 50 kb long nonoverlapping segments. (B) Gene content. The gene content was calculated for 50 kb long segments for both annotated chicken mRNA genes (all_mrna) and Unigene clusters.

CONCLUSION: NONRANDOM RSV-C INTEGRATION AND COMPARISON WITH OTHER RETROVIRUSES

Here we report the first study of RSV integrations in the chicken genome. At the macrolevel, we found that RSV has a preference for large chromosomes, GC-rich and gene-rich regions, and broadly expressed genes. At the microlevel, we have not found any significant preference of RSV integration in the vicinity of transcription start sites and/or promoter

regions. Obviously, some of these features are interdependent. In both mammals and birds, gene-rich regions tend to be GC-rich [25-27]. However, the apparent under-representation of insertions into microchromosomes is surprising, as chicken microchromosomes have, in general, an increased GC content, more CpG islands, and higher gene density

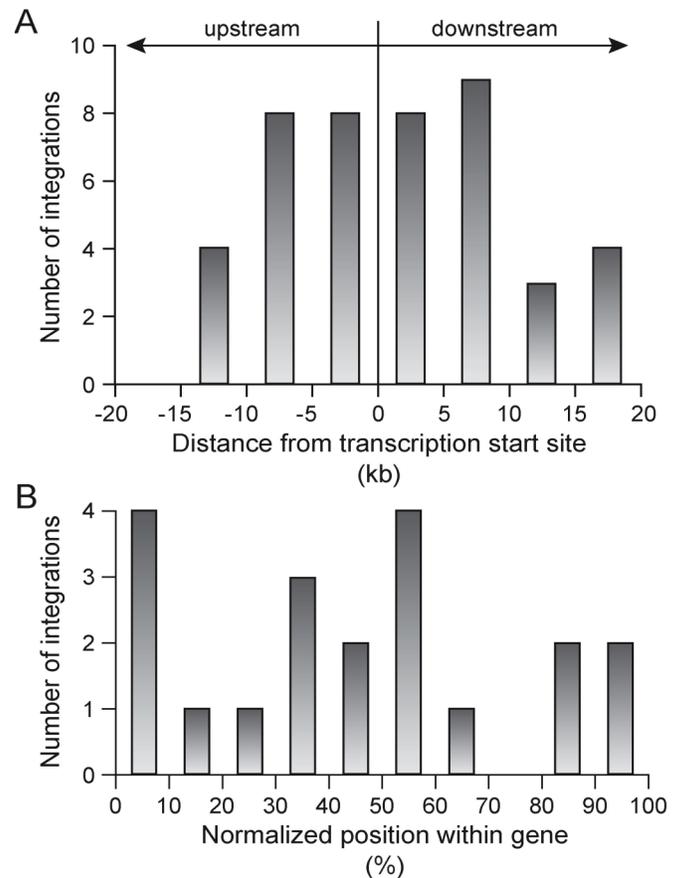


Fig. (4). Distribution of 57 integration sites around cellular genes. To avoid potential artifacts caused by selective removal of redundant records, all targeted genes are counted, even if one insertion was found within or near several transcription units. (A) Distribution of integration sites around transcription starts of annotated nonredundant chicken transcripts obtained from mRNA genes and Unigene clusters. (B) Distribution of integrations within genes. The genes were normalized to a common length to facilitate the comparison.

Table 1. Results of the Expression Analysis for Genes Targeted by Integration and Genes in 100 kb Around the Integration Site

	Genes Targeted by Integration		Genes within 100 kb of the Integration Site		All Genes from UniGene	
	Count	%	Count ²	%	Count	%
Housekeeping genes (expressed in 5 or more tissues)	9	42.9	65***	39.6	8333	27.7
Tissue-specific genes (expressed exclusively in 1 tissue)	3	14.3	30*	18.3	7746	25.7
Preferentially expressed genes ¹ (expressed in 1 or 2 tissues)	5	23.8	53**	32.3	12668	42.1
Total	21	100	164	100	30093	100

¹Including tissue-specific genes.

²Significant difference in comparison to the rest of genes in UniGene.

* $p = 0.03$; ** $p = 0.01$; *** $p = 0.001$ (Fisher's exact test).

[28,29]. The explanation for this observation is unclear but it is in accordance with Barr *et al.* [30] who described the same trend in a bigger dataset of ASLV integration sites. In the same study, however, HIV-1 proviruses accumulated in the microchromosomes.

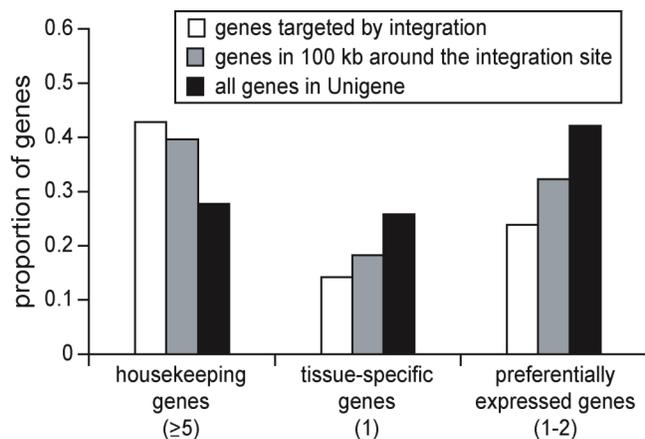


Fig. (5). Proportion of genes targeted by integrations and located in a 100 kb vicinity of integration sites in the context of their expression breadth. The genes were sorted into three categories according to the number of tissues with EST data present (indicated in brackets). The genes within 100 kb of the integration sites displayed significant differences in all three categories in comparison to the rest of the genes in UniGene (see Table 1).

Very interesting is the bias for integration near potential housekeeping genes, but not tissue-specific genes. Given the constitutive expression in all tissues, regions rich in housekeeping genes are expected to have an open chromatin structure, which is favorable for retroviral integration [37]. Interestingly, constitutively expressed genes in both mammals and birds are preferentially found in GC-rich regions in the nuclear interior [38,39]. This may indicate that the higher order chromatin structure is an important factor affecting RSV integration.

The aforementioned integration preferences of RSV are very similar to previously analyzed retroviruses [8,9,13,24,30], despite the fact that some chicken chromosomal and genomic features are different from those of mammalian genomes. In contrast to previous studies on MLV and MLV-derived retroviral vectors in human and murine genomes, we have found that RSV integrates without any significant preference for the transcriptional starts of targeted genes. This resembles the integration of ASLV-derived vectors of the RCAS type (<http://home.ncifcrf.gov/hivdrp/RCAS/plasmid.html>) [13,24,30]. It is tempting to speculate whether the peculiar behavior of MLV in its homologous host genome could be observed also in the heterologous host genome of chicken. Lack of such integration bias would suggest the tethering of MLV PIC with a mammalian-specific DNA binding factor absent in chicken cells. Further genome-wide integration studies in chicken using retroviruses of mammalian origin in comparison with ASLVs are greatly needed for better understanding of this phenomenon.

ABBREVIATIONS

ASLV = Avian sarcoma and leukosis virus

CEF = Chicken embryo fibroblast
 EST = Expressed sequence tag
 HIV-1 = Human immunodeficiency virus type 1
 I-PCR = Inverse polymerase chain reaction
 kb = Kilo bases
 Mb = Mega bases
 MLV = Murine leukemia virus
 PIC = Preintegration complex
 RSV = Rous sarcoma virus
 TU = Transcription unit

AUTHORS' CONTRIBUTIONS

JH designed this study and edited the manuscript. MR elaborated the I-PCR system, cloned integration sites and identified the genome targets. AP, PD, and JH analyzed the genomic distribution of integration sites and the expression of targeted loci. JP, JG and JH prepared the virus and infected DF-1 cells *in vitro*. JH, AP, and PD wrote the manuscript and all authors read and approved its final version.

ACKNOWLEDGEMENTS

The authors are deeply indebted to Daniel Elleder for helpful comments on I-PCR, to Petr Pajer for his help with DNA sequencing, and to Andrew Gentles for critical reading of the manuscript. We also thank Věra Hoserová and Iva Rybková, who performed a lot of the routine cloning work. This research was supported by a grant No. 204/07/1030 to J.P. from the Grant Agency of the Czech Republic and project AV0Z50520514 awarded by the Academy of Sciences of the Czech Republic.

ADDITIONAL INFORMATION

Supplementary data file 1. List of 164 chicken genes located in 100 kb around the integration (see the attached file).

Supplementary data file 2. siteList of analyzed EST libraries (see the attached file).

REFERENCES

- [1] Engelman A. Most of the avian genome appears available for retroviral DNA integration. *BioEssays* 1994; 16: 797-799.
- [2] Holmes-Son ML, Appa RS, Chow SA. Molecular genetics and target site specificity of retroviral integration. *Adv Genet* 2001; 43: 33-69.
- [3] Bushman FD. Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 2003; 115: 135-138.
- [4] Yung E, Sorin M, Wang E-J, Perumal S, Ott D, Kalpana GV. Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance. *J Virol* 2004; 78: 2222-2231.
- [5] Cherepanov P, Devroe E, Silver PA, Engelman A. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem* 2004; 279: 48883-48892.
- [6] Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, *et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 2005; 11: 1287-1289.
- [7] Carteau S, Hoffmann C, Bushman F. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centro-

- meric aliphoid repeats are a disfavored target. *J Virol* 1998; 72: 4005-4014.
- [8] Elleder D, Pavlíček A, Pačes J, Hejnar J. Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. *FEBS Lett* 2002; 517: 285-286.
- [9] Schroeder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002; 110: 521-529.
- [10] Bernardi G. The distribution of proviruses in the mammalian genome. In *Structural and evolutionary genomics. Nature selection in genome evolution*. Edited by Bernardi G. Amsterdam: Elsevier; 2004: 149-160.
- [11] Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 2003; 300: 1749-1751.
- [12] Hematti P, Hong B-K, Ferguson C, *et al.* Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol* 2004; 2: 2183-2190.
- [13] Mitchell RS, Beitzel BF, Schroder ARW, *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2004; 2: 1-11.
- [14] Hacein-Bey-Abina S, Von Kalle C, Schmidt M, *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003; 302: 415-419.
- [15] Mack KD, Jin X, Yu S, *et al.* HIV insertions within and proximal to host cell genes are a common finding in tissues containing high levels of HIV DNA and macrophage-associated p24 antigen expression. *J Acquir Immune Defic Syndr* 2003; 33: 308-20.
- [16] Tsyba L, Rynditch AV, Boeri E, Jabbari K, Bernardi G. Distribution of HIV-1 in the genomes of AIDS patients. *Cell Mol Life Sci* 2004; 61: 721-726.
- [17] Han Y, Lassen K, Monie D, *et al.* 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* 2004; 78: 6122-6133.
- [18] Hanai S, Nitta T, Shoda M, *et al.* Integration of human T-cell leukemia virus type 1 in genes of leukemia cells of patients with adult T-cell leukemia. *Proc Natl Acad Sci USA* 2004; 95: 306-310.
- [19] Ozawa T, Itoyama T, Sadamori N, *et al.* Rapid isolation of viral integration site reveals frequent integration of HTLV-1 into expressed loci. *J Hum Genet* 2004; 49: 154-165.
- [20] Taganov KD, Cuesta I, Daniel R, *et al.* Integrase-specific enhancement and suppression of retroviral DNA integration by compacted chromatin structure *in vitro*. *J Virol* 2004; 78: 5848-5855.
- [21] Hejnar J, Plachý J, Geryk J, *et al.* Inhibition of the Rous sarcoma virus long terminal repeat-driven transcription by *in vitro* methylation: Different sensitivity in permissive chicken cells versus mammalian cells. *Virology* 1999; 255: 171-181.
- [22] Hejnar J, Hájková P, Plachý J, Elleder D, Stepanets V, Svoboda J. CpG island protects Rous sarcoma virus-derived vectors integrated into nonpermissive cells from DNA methylation and transcriptional suppression. *Proc Natl Acad Sci USA* 2001; 98: 565-569.
- [23] Hejnar J, Elleder D, Hájková P, Walter J, Blažková J, Svoboda J. Demethylation of host-cell DNA at the site of avian retrovirus integration. *Biochem Biophys Res Commun* 2003; 311: 641-648.
- [24] Narezkina A, Taganov KD, Litwin S, *et al.* Genome-wide analyses of avian sarcoma virus integration sites. *J Virol* 2004; 78: 11656-11663.
- [25] Cortadas J, Olofsson B, Meunier-Rotival M, Macaya G, Bernardi G. The DNA components of the chicken genome. *Eur J Biochem* 1979; 99: 179-186.
- [26] Olofsson B, Bernardi G. Organization of nucleotide sequences in the chicken genome. *Eur J Biochem* 1983; 130: 241-245.
- [27] Bernardi G. Isochores and the evolutionary genomics of vertebrates. *Gene* 2000; 241: 3-17.
- [28] Andreozzi L, Federico C, Motta S, *et al.* Compositional mapping of chicken chromosomes and identification of the gene-richest regions. *Chromosome Res* 2001; 9: 521-532.
- [29] International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 2004; 432: 695-716.
- [30] Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD. Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J Virol* 2005; 79: 12035-12044.
- [31] Gudkov AV, Obukh IB, Serov SM, Naroditsky BS. Variety of endogenous proviruses in the genomes of chickens of different breeds. *J Gen Virol* 1981; 57: 85-94.
- [32] Méric C, Spahr P-F. Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J Virol* 1986; 60: 450-459.
- [33] Kent WJ. BLAT – the BLAST-like alignment tool. *Genome Res* 2002; 12: 656-664.
- [34] Karolchik D, Baertsch R, Diekhans M, *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res* 2003; 31: 51-54.
- [35] Boguski MS, Lowe TM, Tolstoshev CM. dbEST – database for expressed sequence tags. *Nat Genet* 1993; 4: 332-333.
- [36] Wheeler DL, Barrett T, Benson DA, *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2005; 33: D39-45.
- [37] Vijaya S, Steffen DL, Robinson HL. Acceptor sites for retroviral integrations map near Dnase I-hypersensitive sites in chromatin. *J Virol* 1986; 60: 683-692.
- [38] Saccone S, Federico C, Bernardi G. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. *Gene* 2002; 300: 169-78.
- [39] Lercher MJ, Urrutia AO, Pavlicek A, Hurst LD. A unification of mosaic structures in the human genome. *Hum Mol Genet* 2003; 12: 2411-2415.

Received: January 15, 2008

Revised: February 12, 2008

Accepted: February 13, 2008

© Reinišová *et al.*; Licensee *Bentham Open*.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.