RESEARCH PAPER

Mobile Genetic Elements 2:3, 133-141; May/June, 2012; © 2012 Landes Bioscience

Distribution of MGEs and their insertion sites in the *Macaca mulatta* genome

Kamal Rawal,[1,†,*] Avantika Priya,[1,†] Aman Malik,[1] Radhika Bahl and Ram Ramaswamy[2]

[1]Department of Biotechnology; Jaypee Institute of Information Technology; Noida, India; [2]University of Hyderabad; Central University; Gachibowli, Hyderaba, India

[†]These authors contributed equally to this work.

Keywords: mobile genetic elements, primates, LINEs, SINEs, Alu, L1, truncation points, physicochemical properties

Mobile genetic elements (MGEs) are fragments of DNA that can move around within the genome through retrotransposition. These are responsible for various important events such as gene inactivation, transduction, regulation of gene expression and genome expansion. The present work involves the identification and study of the distribution of Alu and L1 retrotransposons in the genome of *Macaca mulatta,* an extensively used organism in biomedical studies. We also make comparisons with MGE distributions in other primate genomes and study the physicochemical properties of the local DNA structure around the transposon insertion site using ELAN. The present work also includes computational testing of the pre-insertion loci in order to detect unique features based on DNA structure, thermodynamic considerations and protein interaction measures. Although there is significant sequence divergence between the elements of *M. mulatta* and *H. sapiens,* their genome wide distribution is very similar; comparing the distribution of L1's in all available X chromosome sequences suggests a common mechanism behind the spread of MGE's in primate genomes.

Introduction

The mammalian order primate that includes humans, apes and monkeys in addition to several other organisms can be traced to the late Cretaceous period. The rhesus macaque is in many ways an ideal model organism, being closely related to humans (sharing a common ancestor about 25 million years ago) and also sharing similar physiology, neurobiology and susceptibility to infectious and metabolic diseases. Since both the *M. mulatta* and *H. sapiens* genomes have been sequenced, it is known that the evolutionary distance between them is small, with local fluctuations and low divergence, particularly in chromosome X. On average, orthologs have about 97% identity between the genomes both at the nucleotide and amino acid sequence levels. Approximately 50% of the rhesus macaque genome consists of various repetitive sequences, similar to the human genome.[1-3]

The phenomenon of retrotransposition that occurs in eukaryotic genomes of diverse taxonomic groups is implicated in various human genetic diseases. Insertion sites of many non-long-terminal repeat (LTR) retrotransposons play an important role in genome evolution and are distributed throughout the genome. The phenomenon behind the selection of the insertion sites of these elements has been shown to be correlated with patterns found at pre-insertion loci.[4] It is well known that mobile elements insertions are capable of altering gene expression,[5] generating genomic deletions,[6] and they can even create new genes and gene families.[7] Existing repetitive elements can also cause ectopic recombinations.[8]

Despite the overall similarity in retrotransposon mobilization activity in the old world monkeys and hominid lineages, mobile elements have continued to evolve independently in both lineages. The retroelements that are presumed to have had the most dramatic impact in shaping primate genomes are the L1 family of LINE elements and the Alu elements, their partner SINEs.[9] Besides contributing large amounts of DNA to many genomes (including at least 40% of the human genome) they have also provided new genes, exons and other motifs involved in the physical and sequence structure of chromosomes. There are instances in which a previous element is now a part of the machinery that regulates gene expression.[10]

Repetitive elements account for about 50% of the genome[3] among all of the presently sequenced primate species (**Table 1**). In *M. mulatta,* two classes of mobile elements are present, class I DNA transposons and class II retrotransposons. The transposons can also be categorized into different families and subfamilies, based on the relationships between their sequences. The rhesus family consists of about 320,000 copies of many families of DNA transposons and about half a million copies of endogenous retroviruses. The L1s and Alu elements account for most of the lineage specific insertions and these have been playing an important role in shaping the complete genome.[3,11]

*Correspondence to: Kamal Rawal; Email: kamal.rawal@jiit.ac.in
Submitted: 04/19/12; Revised: 06/07/12; Accepted: 06/08/12
http://dx.doi.org/10.4161/mge.21074

Table 1. Summary of the number of L1 and Alu elements present in Macaca genome in comparison with Human and *Pan troglodytes*[3,11]

Species	LINE(L1)	SINE (Alu)
Human	104,541	1,144,000
Pan	558,000	1,111,000
Rhesus	100,000	1,076,800

Table 2A. Table showing the percentage similarity between the L1 of human and Macaca genome

	L1	L1P4a	L1P4b	L1P4c	L1P4d	L1P4e
L1	100%	25.80%	20%	18.10%	23.60%	13.30%
L1P4a	25.80%	100%	41.30%	38.30%	42.70%	38.50%
L1P4b	20%	41.30%	100%	52.50%	44.30%	44.20%
L1P4c	18.10%	38.30%	52.50%	100%	46.70%	45.70%
L1P4d	23.60%	42.70%	44.30%	46.70%	100%	55.20%
L1P4e	13.10%	38.50%	44.20%	45.70%	55.20%	100%

Table 2B. Table showing the percentage similarity between the Alu of human and Macaca genome

	Alu	AluMcaYa2	AluMcaYa3	AluMcaYb4
Alu	100%	76.80%	79.30%	75.70%
AluMcaYa2	76.80%	100%	91.90%	96.70%
AluMcaYa3	79.30%	91.90%	100%	90.90%
AluMcaYb4	75.70%	96.70%	90.90%	100%

ELAN was developed earlier as a suite of tools for genome-wide retrotransposon element analysis.[12] The application of modules of this bioinformatics pipeline is described as follows:

(1) ELEFINDER performs a whole genome distribution analysis of the MGEs through a BLASTN search by making the use of Perl/BioPerl scripts by which the output files are parsed. It also extracts sequence 100 bp up and downstream at each MGE site identified as a *preinsertion* locus.[12]

(2) DNASCANNER scans DNA sequences such as preinsertion loci and analyses insertion hotspots of elements in detail so as to provide a set of signals or characteristics that are potentially recognized by an element for its insertion.[12]

The *M. mulatta* genome has a total of 22 chromosomes including X and Y (although the sequencing project did not sequence Y). The only Y chromosome that has been completely sequenced is of humans,[14,15] and sequencing of the chimpanzee and mouse Y chromosomes is in progress.[15-17] The sequencing of the mammalian Y chromosomes, of the organisms like rhesus macaque (*Macaca mulatta*), the white-tufted-ear marmoset (*Callithrix jacchus*), the rat (*Rattus novergicus*), the bull (*Bos taurus*) and the opossum (*Monodelphis domestica*), is proposed and still under process.[17]

We present here a study of the primate *Macaca mulatta* genome to identify and characterize insertion sites of the two representative retroelements present, and further, comparison with similar features of the human genome (excluding the Y chromosome). The structural and thermodynamic features as well as protein interaction measures are computed in preinsertion loci using the tool DNA Scanner.

Results

In the human genome, a full-length L1 element is around 6 kb long, and is reported to be the most successful TEs in human genome by mass while Alu elements, typically ~300 bp long[18] are most successful in terms of copy number. Currently, there are three macaque consensus sequences for Alu: AluMacYa3, AluMacYb2, and AluMAcYb4 and five of that for L1: L1P4a, L1P4b, L1P4c, L1P4d, L1P4e in Repbase (Version 13.5).[19]

We first compute the nucleotide sequence divergence of L1 (human) with respect to the corresponding L1 (Macaca genome) (**Table 2A**). As can be seen, pairwise alignment shows a low percentage identity, with the major areas of dissimilarity in the 3' region. The macaca lineages show higher percentage similarity with each other; with regard to the human genome, the L1 nearest to its human analog is L1P4d, while that which has diverged the most is L1P4e.

Macaca lineages show high similarity with each other and human Alu seems to be most closely related to the Macaca specific AluMcaYa3 (**Table 2B**). There are approximately 1.1 million ALU and 95,000 L1 copies in the Macaca genome. **Tables 3 and** 4 give details of the numbers of Alu and L1 elements in each chromosome. The four groups constructed for the pre insertion loci were: intact on both ends, intact on 5', intact on 3' and intact on neither end.

About 1077410 copies of Alus were found uniformly distributed on each chromosome. Of these approximately 13.49% (145,428) are truncated on the 5' end, 456,752 copies (42.39%) are truncated at both the ends, 30.29% are the truncated at 3' end (326,447). Only 148,783 are intact at both the ends (13.8%).

A negligible fraction (1.6%) of L1 elements was intact on both sides: of the 94,616 elements identified 22% (21,044) were truncated at the 5' end and 1.38% (1,314) were truncated at the 3' end and about 75% at both ends.

As in the human genome, there are few functional Alu copies in *M. mulatta,* and their distribution on the various chromosomes follows similar patterns (**Fig. 1A**). The X chromosome is known to have an exceptionally high number of the L1 elements (**Fig. 1B**), which also contained the largest number of truncated elements (at both the ends).

The pre-insertion loci were extracted and evaluated for various physicochemical properties as described in our earlier paper.[12] The positions of the extrema were similar to those seen in other cases in the physicochemical profiles generated by DNASCANNER.

Similar tables were calculated for each *M. mulatta* chromosome and for all the 14 characteristics extrema were seen in the range of -9 to -11 bp for Alu element and that for the L1 element was between -2 to -19 for the majority of the cases.

For L1 elements, results are similar and **Table 5** gives the complete information about the values and the extrema for each of the properties.

Figures 2 and 3 show the graphs obtained for four physicochemical properties for Alu and L1 elements insertion sites in

Table 3. Alu element distribution on the different chromosomes on the *M. mulatta* genome

Chromosome no.	TOTAL	Truncated at both ends	5' Truncated & 3' intact	5' Intact & 3' truncated	Intact on both ends
1	103699	44769	13593	31424	13913
2	61712	25396	8860	18475	8981
3	71346	30537	9480	21743	9586
4	56574	23427	7909	17041	8197
5	50565	20280	7969	14779	7537
6	54732	22044	8079	16564	8045
7	70113	29889	9379	21225	9620
8	46914	19670	6653	13949	6642
9	52238	22445	7096	15740	6957
10	49717	21500	6107	15674	6436
11	57920	25428	7465	17379	7648
12	33564	13738	4897	10014	4915
13	46262	18827	6283	14224	6928
14	44333	18676	5954	13483	6220
15	42923	17631	5765	13445	6082
16	51097	22740	5794	16036	6527
17	26701	10529	3966	8080	4126
18	22914	9211	3222	7099	3382
19	47666	23303	5244	13934	5185
20	43127	20163	5249	12809	4906
X	43293	16549	6464	13330	6950
Total	1077410	456752	145428	326447	148783

Table 4. L1 element distribution on the different chromosomes on the macaca genome

Chromosome no.	TOTAL	Truncated at both ends	5' Truncated & 3' intact	5' Intact & 3' truncated	Intact on both ends
1	6918	5234	1477	96	111
2	7056	5271	1583	87	115
3	5791	4409	1213	84	85
4	6123	4558	1399	88	78
5	7652	5541	1869	120	122
6	6980	5093	1654	106	127
7	5102	3819	1136	60	87
8	4940	3638	1136	69	97
9	3768	2821	837	44	66
10	1819	1461	321	13	24
11	4194	3080	976	61	77
12	3489	2549	814	63	63
13	4352	3274	938	67	73
14	4167	3075	961	67	64
15	3417	2609	724	35	49
16	1257	969	258	21	9
17	3517	2605	821	46	45
18	2305	1738	505	26	36
19	889	701	159	19	10
20	1276	993	253	12	18
X	9604	7320	2010	130	144
Total	94616	70758	21044	1314	1500

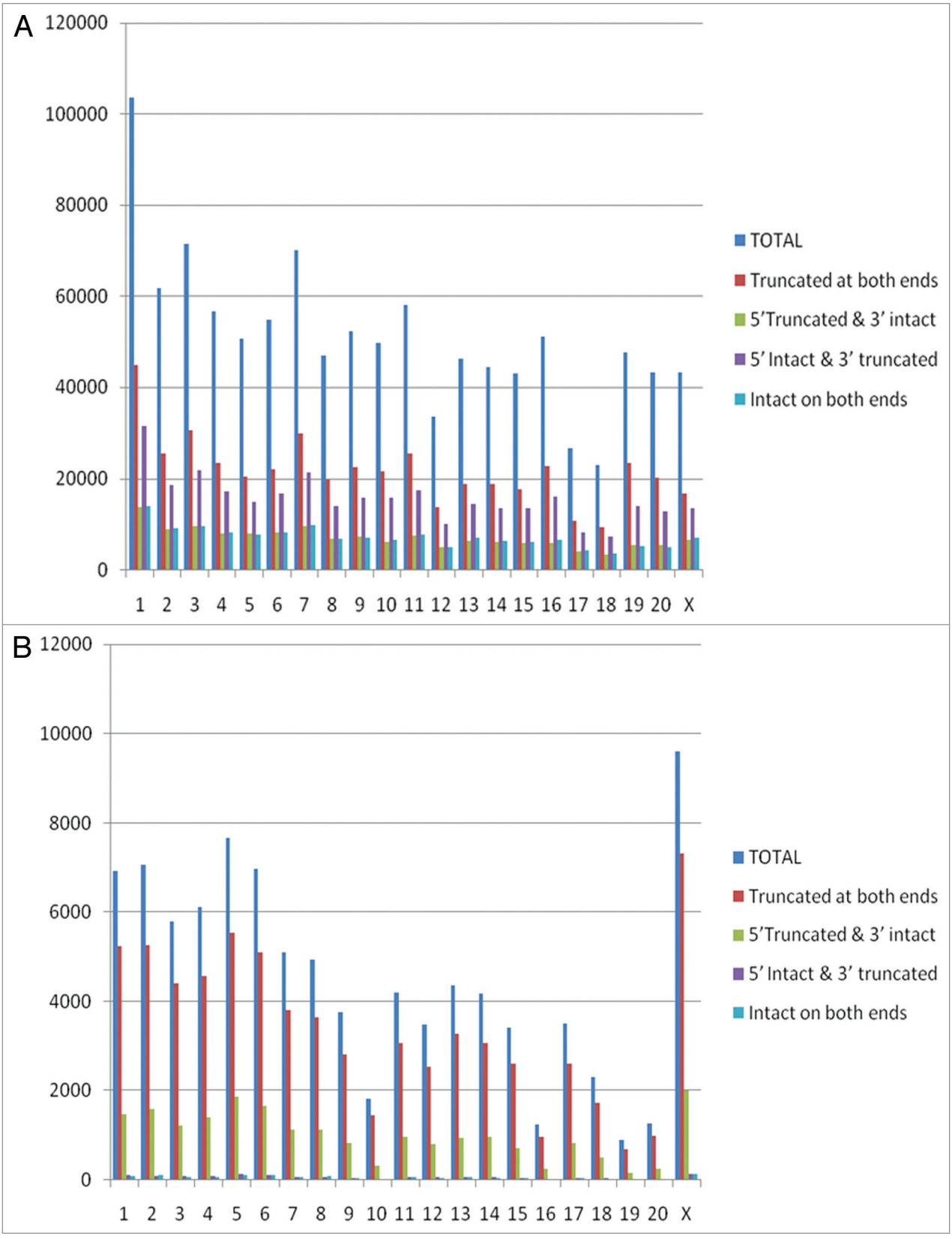

Figure 1. (A) Distribution of Alu element across macaca genome. Four classes of elements are indicated with four different colors. The y-axis represents the frequency of elements found on the different chromosomes (marked along the x-axis). (B) Distribution of L1 element across macaca genome. The y-axis represents the frequency of elements found on the different chromosomes (marked along the X-axis).

Table 5. Information derived from DNASCANNER analysis of the *M. mulatta* chromosome 1 for Alu and L1 elements

	SINEs (Alu)			LINEs (L1)		
PROPERTIES	**Trend**	**Position**	**Value**	**Trend**	**Position**	**Value**
A rule	U	-9	0.480	U	-16	0.528
AT rule	U	-10	0.743	U	-19	0.771
b-a-trimeric1	U	-10	0.280	U	-17	0.288
bend_scl1	D	-9	-0.021	D	-14	-0.026
bendingstiffness1	D	-10	23.457	D	-19	22.463
C rule	D	-9	0.104	D	-14	0.104
duplexstability-freeenergy	U	-10	-0.674	U	-19	-0.659
G rule	D	-11	0.145	D	-19	0.116
np_scl1	D	-10	-3.583	D	-18	-4.197
propellartwist1	D	-10	-7.370	D	-18	-7.507
proteininducedform1	D	-10	1.966	D	-18	1.923
stabilizingenergy_zdna1	U	-10	1.787	U	-14	1.814
stackingenergy1	U	-10	-3.363	U	-19	-3.289
T rule	D	-5	0.222	D	-2	0.191

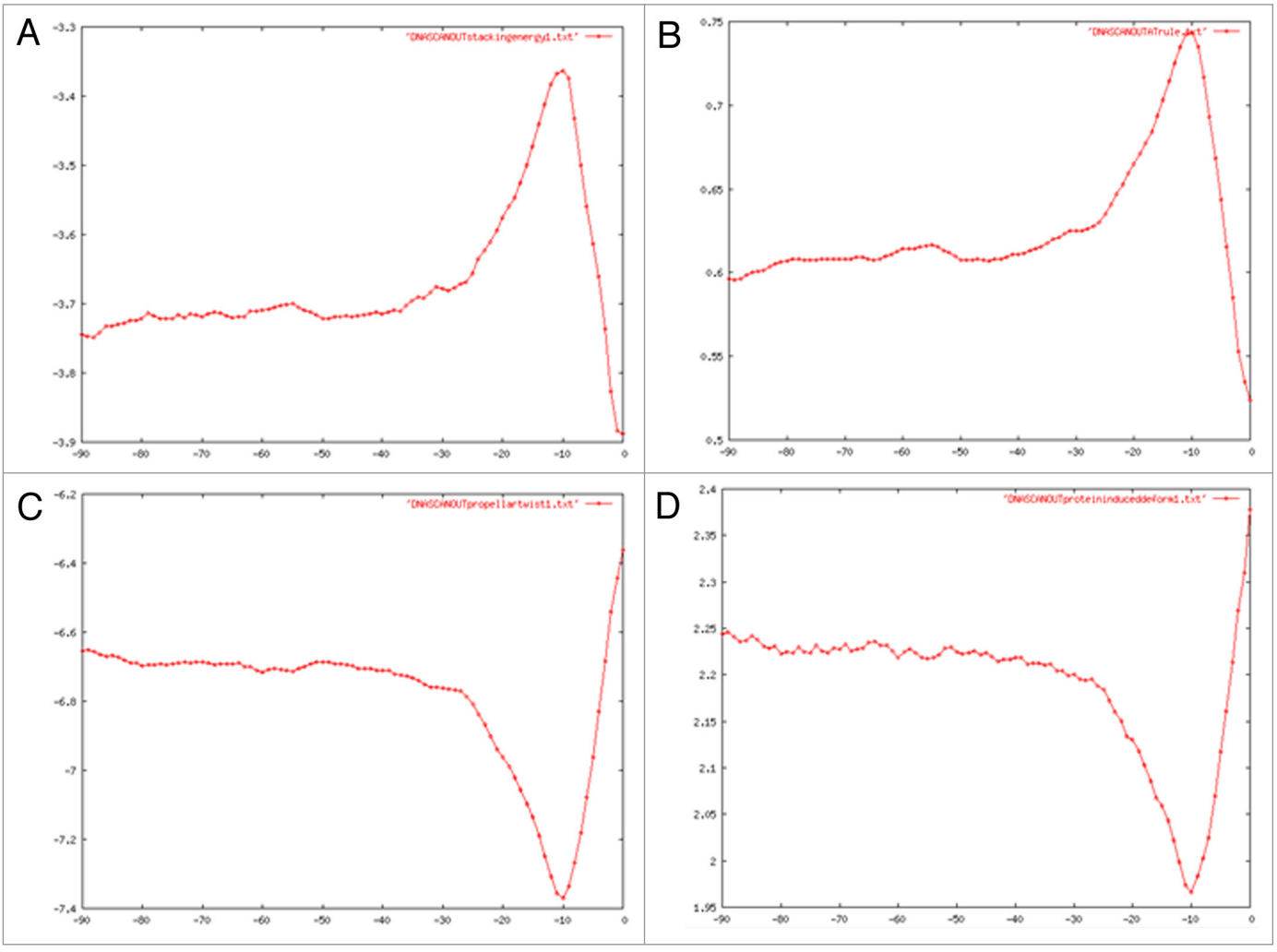

Figure 2. Various signals upstream of the insertion sites of Alu in chromosome 1, for (A) stacking energy, (B) AT content, (C) Propeller twist and (D) protein induced deformability. The y axis represents value of the property and the x-axis gives the relative position with respect to the insertion site (taken to be 0).

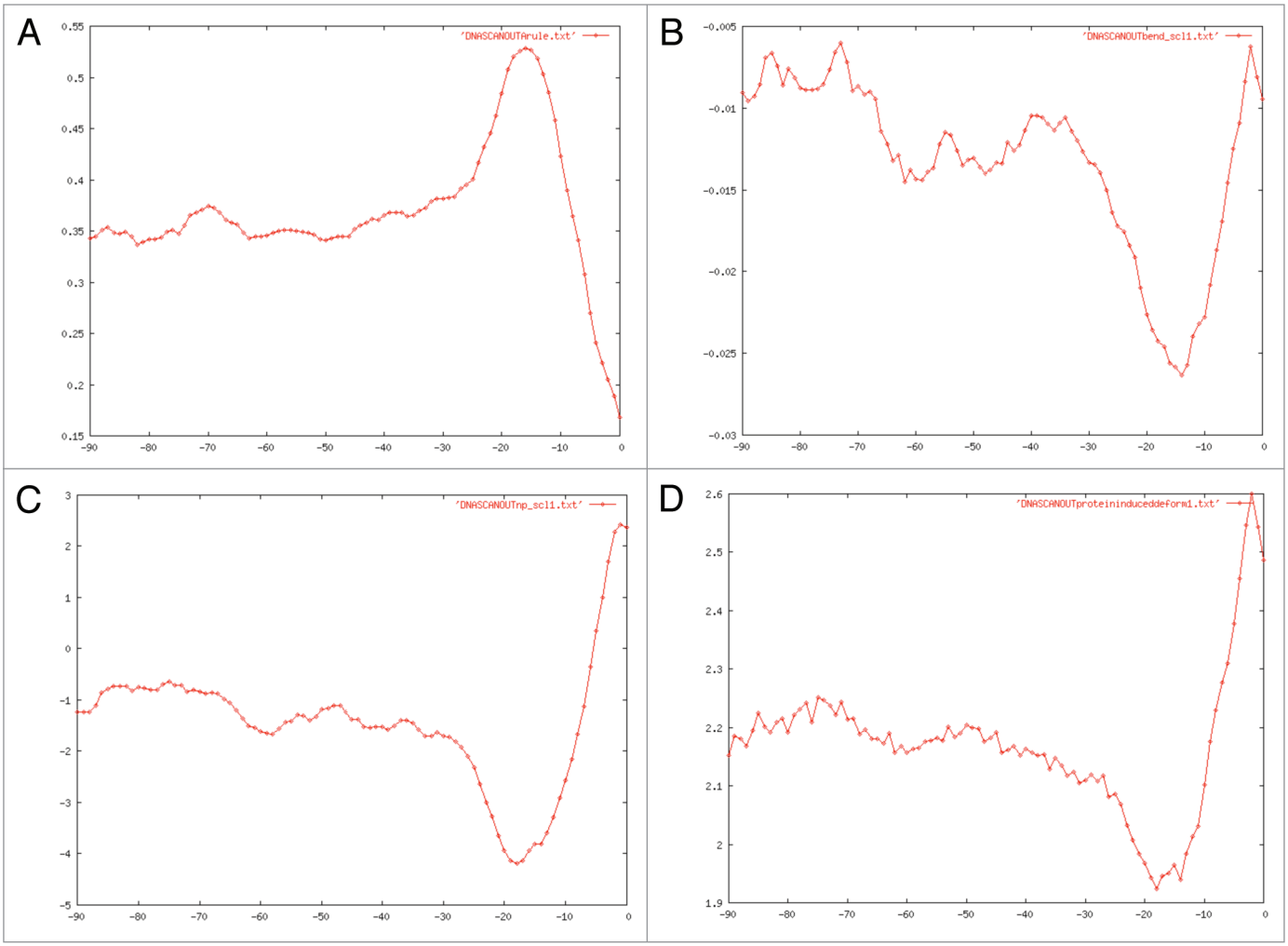

Figure 3. Various signals upstream of the insertion sites of L1 in chromosome 1, for (A) A rule, (B) nucleosomal bending, (C) nucleosomal positioning, (D) protein induced deformability. The y axis represents value of the property and the x-axis gives the relative position with respect to the insertion site (taken to be 0).

macaca genome. The trend followed is the same for both elements. Control sequences were generated by scrambling the positive data set of pre-insertion sequences; all these above properties gave a featureless distribution (namely no extrema). Another independent set of control sequences that were obtained by randomly selecting genomic sequences of 100 bp also gave similar featureless results.

Our rationale for the choice of the various parameters are briefly given below:

(1) Regions with alternating purines/pyrimidines steps and AT rich regions melt more readily. Therefore DNA denaturation profiles were computed for insertion sites. We found regions of low GC and high AT content, indicating that a relatively less energy is required to melt DNA near insertion sites, which in turn favors retrotransposition (**Figs. 2B and 3A**).

(2) Propeller twist is a property, involved in the distortion of the hydrogen bonds that hold two bases together. Regions with specific dinucleotides with large propeller twist, followed by a lower propeller twist were obtained (**Fig. 2C**), which shows that latter regions may be easily distorted and are suitable for insertion.

(3) Nucleosomes are involved in DNA compacting and providing transcription factors access to the respective regulatory regions. Two different nucleosomal related features, the bending energy/persistence length and the position profiles of the nucleosomes, were studied.[21] Regions with comparatively low energies were obtained, within the upstream areas of the insertion sites. (**Fig. 3B and C**).

(4) Stacking energy profiles showed a maximum near -10, indicating that this region is unstable, leading to easy de-stacking of DNA sequence, which would thereby enable an easy insertion of Alu. (**Fig. 2A**).

(5) Duplex stability is a measure of the relative stability of the DNA-duplex structure, which is directly dependent on sequence.[21] We obtained the region around the -10 position for Alu and at -19 for L1, with a peak, representing a region which would de-stack or melt easily.

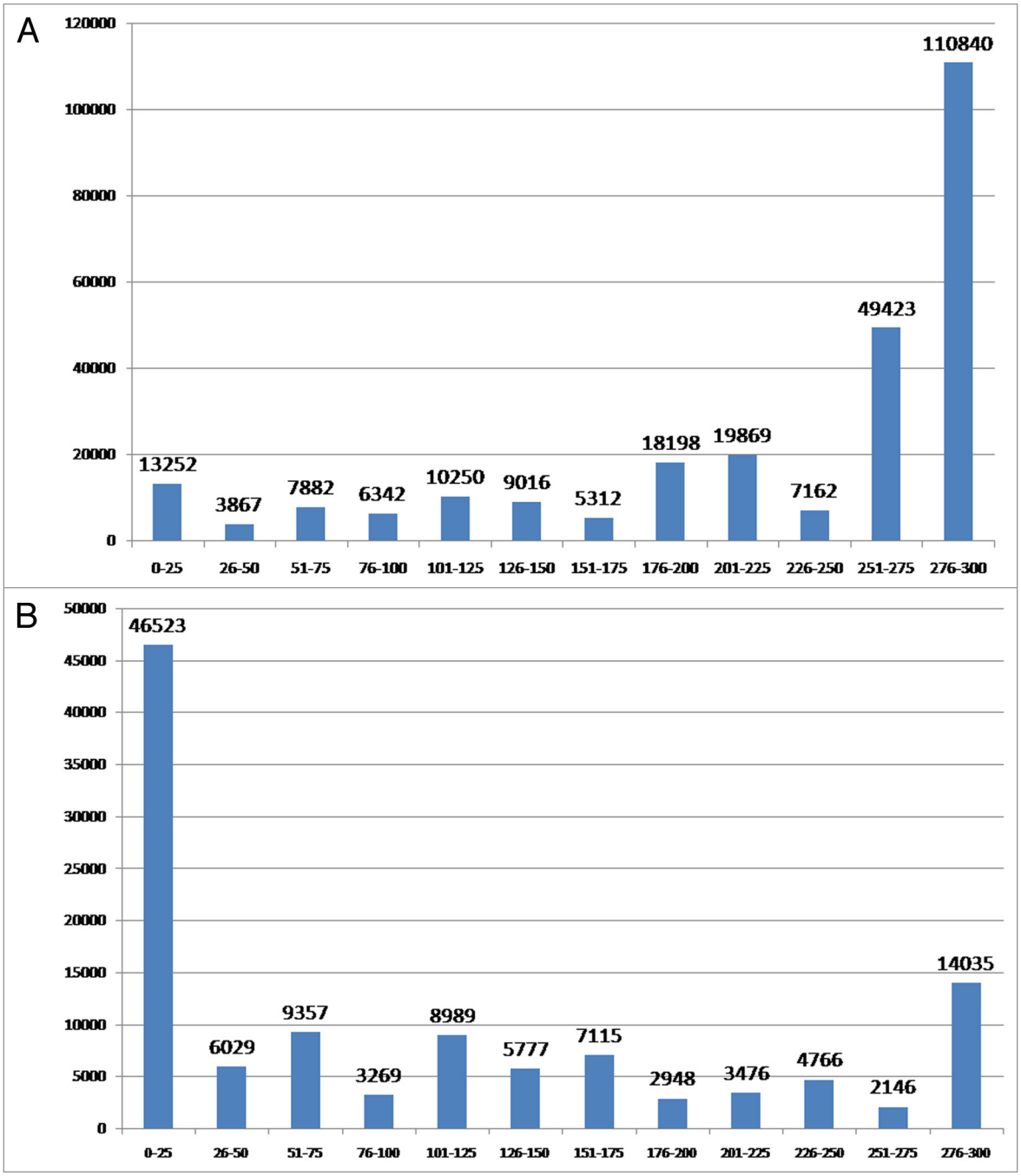

Figure 4. (A) The truncation distribution of Alu element in *Macaca mulatta* for bin size 25 and the graph is plotted for 3' end of Alu truncated and its 5' end being intact. (B) The truncation distribution of Alu element in *Macaca mulatta* for bin size 25 and the graph is plotted for 5' end of Alu truncated and its 3' end being intact.

(6) DNA deformability is an important property, dependent on the sequence and required for interaction with proteins. The DNA deformability was calculated and a region (at -10 for Alu and -18 for L1) of low deformability was seen; this facilitates retrotransposon insertion. (**Figs. 2D and 3D**).

The results of ELEFINDER as well as the information curated in the InSiDe database were used to find the distribution of truncation sites in the whole *M. mulatta* genome. The x-axis represents the length of element divided in bins of 25, i.e.,1–25, 25–50 and so on. The criteria used for the plotting was the occurrence of 5'truncated and 3' intact ends and 3' truncated and 5' intact ends in the Macaca genome. The basic aim was to identify the positions in the element sequence, where most of the times truncation occurs for the macaca genome as a whole.

The graph plotted for the 3' truncated ends, the maximum number of hits were obtained in the last bin, i.e., the 276–300 bin has the maximum number of truncations in the Macaca

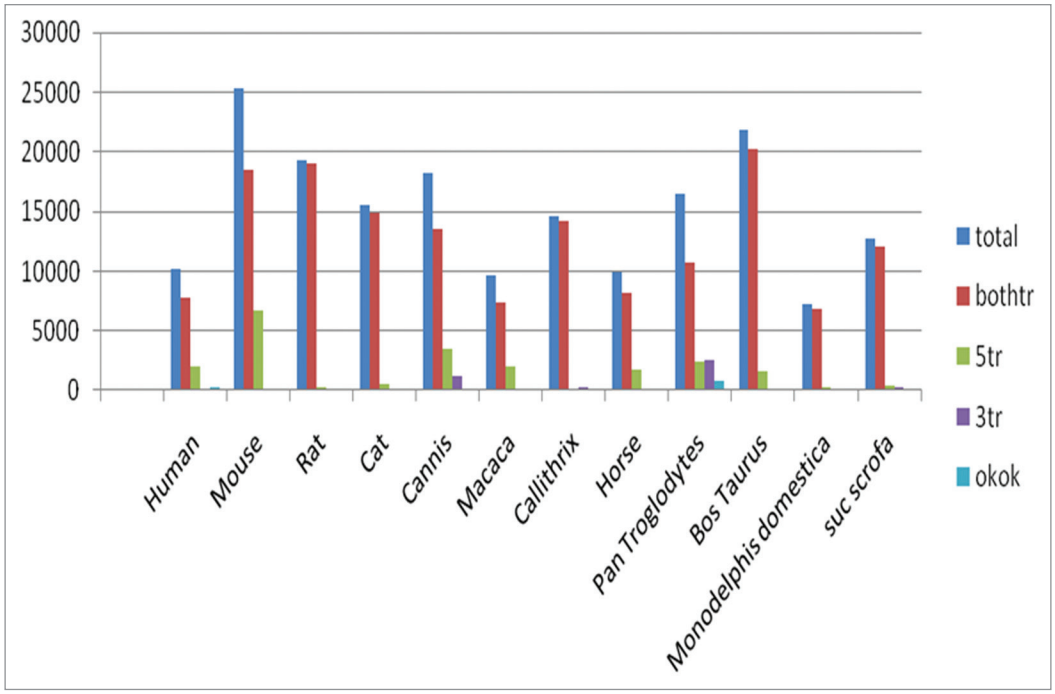

Figure 5. L1 element distribution on the X-chromosome of different organisms.

genome (**Fig. 4A**). Similarly, for the 5' truncated category, the maximum number of truncations was found to be in the first bin, i.e., the 1–25 bin (**Fig. 4B**).

Discussion

Previously, we analyzed several genomes for distribution of MGEs as well as their insertion sites along with the signals facilitating their insertion.[12] In the present study of Alus (SINE) and L1s (LINE) in the Macaca genome we found that the insertion sites had physicochemical characteristics that were similar to those observed in other organisms, suggesting that these are generally important. The present study confirmed that the region 100 bp upstream from Alu and L1 insertion sites show statistically significant distinctive properties both in the physical and structural characteristics, as well as in the energetics. These properties seem to play an important role in the insertion of MGE. During insertion, a MGE causes the target site to distort in a number of ways and requires the cooperative action of a number of proteins to break bonds, unwind the DNA and to nick the target site strand.[13] It is due to this series of requirements, that the insertion sites for all the chromosomes, show a characteristic set of physicochemical properties, signified by the extremum peaks in each case.

Each of the peaks for the several properties carry biological significance and may be used further for the identification of potential new insertion sites. In each case, that the nature of the extremum that has a role in defining the trend of each property was identical for both Alu and L1 (**Table 5**), explaining that the signals that are needed for the insertion of the element is same, and are probably necessary for the insertion to actually occur, although also probably are not sufficient.

Detection of the most probable truncation sites for Alu elements within the complete Macaca genome revealed that during insertion the truncation distribution of Alu peaks toward the starting positions in the case of the trailing edge truncation and is reversed for leading edge truncations.

We have found that LINEs are present in large number in all the primate genomes which includes the recently studied gorilla genome[22] as well as *Callithrix jaqqus*, *Pan troglodytes*, orangutan (**Fig. 5**). There could be an evolutionary link in primate genomes through the spread of LINEs and SINEs; alternately, there may have been a "master" L1 or retrotransposon copy in the genome of last common ancestor of all primates.

The present work adds insight into primate genome architecture, showing the common structural features that promote MGE insertion and genome expansion. In future work, we aim to compare the insertion sites across species—a task that will be facilitated as the diversity of sequenced genomes increases further.

Methods

The genome sequence of *Macaca mulatta* was retrieved from the NCBI (ftp server: ftp://ftp.ncbi.nih.gov/genomes/). The element sequence (for Alu and L1) were obtained from RepBase[16] and pairwise alignment was performed for the L1 for the human and Macaca specific L1s and its lineages using standard procedures. Copies of L1 and Alu, were mapped on the genome using ELEFINDER.[12]

ELEFINDER was used to find the insertion site of ALU and L1 in the Macaca genome. This tool finds the nature, distribution, genomic location and the site of truncation for each of the insertion sites and performs comparative genome analysis and

also generates several set of sequences. Since we have previously shown that intact copies show the presence of signal as compared with the truncated groups, the analysis of the full-length elements, i.e., of those capable of transposition was also performed.

The tool DNA Scanner analyses the DNA for many physicochemical properties by using various thermodynamic, protein interactions and sequence-based features, which are beyond the T density and AT density. In accordance to the choice of input parameters, the program evaluates a number of properties, using windows that move along the length of the query DNA sequence.[12,23]

Since downstream sequences tend to show signals that are not sufficiently statistically significant only upstream sequences were investigated for both Alu and L1 elements. The controls were selected by scrambling sequences, randomly picking sequences from genome as well as gene sequences.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 2005; 437:69-87; PMID:16136131; http://dx.doi.org/10.1038/nature04072.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11237011; http://dx.doi.org/10.1038/35057062.
3. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al.; Rhesus Macaque Genome Sequencing and Analysis Consortium. Evolutionary and biomedical insights from the rhesus macaque genome. Science 2007; 316:222-34; PMID:17431167; http://dx.doi.org/10.1126/science.1139247.
4. Mandal PK, Rawal K, Ramaswamy R, Bhattacharya A, Bhattacharya S. Identification of insertion hot spots for non-LTR retrotransposons: computational and biochemical application to Entamoeba histolytica. Nucleic Acids Res 2006; 34:5752-63; PMID:17040894; http://dx.doi.org/10.1093/nar/gkl710.
5. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 2003; 13:651-8; PMID:14638329; http://dx.doi.org/10.1016/j.gde.2003.10.013.
6. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, et al. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res 2005; 33:4040-52; PMID:16034026; http://dx.doi.org/10.1093/nar/gki718.
7. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A 2006; 103:17608-13; PMID:17101974; http://dx.doi.org/10.1073/pnas.0603224103.
8. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002; 3:370-9; PMID:11988762; http://dx.doi.org/10.1038/nrg798.
9. Hasnaoui M, Doucet AJ, Meziane O, Gilbert N. Ancient repeat sequence derived from U6 snRNA in primate genomes. Gene 2009; 448:139-44; PMID:19647053; http://dx.doi.org/10.1016/j.gene.2009.07.015.
10. Vincent BJ, Myers JS, Ho HJ, Kilroy GE, Walker JA, Watkins WS, et al. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol Biol Evol 2003; 20:1338-48; PMID:12777507; http://dx.doi.org/10.1093/molbev/msg146.
11. Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE. Comparative analysis of Alu repeats in primate genomes. Genome Res 2009; 19:876-85; PMID:19411604; http://dx.doi.org/10.1101/gr.083972.108.
12. Rawal K, Ramaswamy R. Genome-wide analysis of mobile genetic element insertion sites. Nucleic Acids Res 2011; 39:6864-78; PMID:21609951; http://dx.doi.org/10.1093/nar/gkr337.
13. Meysman P, Marchal K, Engelen K. DNA structural Properties in functional. Genomics 2012; In press.
14. Hughes JF, Skaletsky H, Rozen S, Wilson RK, Page DC. Has the chimpanzee Y chromosome been sequenced? Nat Genet 2006; 38:853-4, author reply 854-5; PMID:16874316; http://dx.doi.org/10.1038/ng0806-853b.
15. Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, et al. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 2005; 437:100-3; PMID:16136134; http://dx.doi.org/10.1038/nature04101.
16. Kuroki Y, Toyoda A, Noguchi H, Taylor TD, Itoh T, Kim DS, et al. Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. Nat Genet 2006; 38:158-67; PMID:16388311; http://dx.doi.org/10.1038/ng1729.
17. Rozen S, Warren WC, Weinstock G, O'Brien SJ, Gibbs RA, Richard K, et al. Sequencing and Annotating New Mammalian Y Chromosomes. A White Paper Proposal. 2006.
18. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009; 10:691-703; PMID:19763152; http://dx.doi.org/10.1038/nrg2640.
19. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 2005; 110:462-7; PMID:16093699; http://dx.doi.org/10.1159/000084979.
20. Schmitz J, Roos C, Zischler H. Primate phylogeny: molecular evidence from retroposons. Cytogenet Genome Res 2005; 108:26-37; PMID:15545713; http://dx.doi.org/10.1159/000080799.
21. Sivolob AV, Khrapunov SN. Translational positioning of nucleosomes on DNA: the role of sequence-dependent isotropic DNA bending stiffness. J Mol Biol 1995; 247:918-31; PMID:7723041; http://dx.doi.org/10.1006/jmbi.1994.0190.
22. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. Insights into hominid evolution from the gorilla genome sequence. Nature 2012; 483:169-75; PMID:22398555; http://dx.doi.org/10.1038/nature10842.
23. Bakre AA, Rawal K, Ramaswamy R, Bhattacharya A, Bhattacharya S. The LINEs and SINEs of Entamoeba histolytica: comparative analysis and genomic distribution. Exp Parasitol 2005; 110:207-13; PMID:15955314; http://dx.doi.org/10.1016/j.exppara.2005.02.009.