Indian Academy of Sciences Conference Series (2019) 2:1
DOI: 10.29195/iascs.02.01.0005

Identification and analysis of mobile-genetic elements in gibbon genome

KAMAL RAWAL[1,2,3,*], JAISRI JAGANNADHAM[1], CHAHAT KUBBA[1] and TANYA SHARMA[1]

[1]Department of Biotechnology, Jaypee Institute of Information Technology, A-10 Sector 62, Noida 201 309, India
[2]Amity Institute of Biotechnology, Amity University, Sector 125, Noida 201 313, India
[3]National School of Tropical Medicine, Baylor College of Medicine, Houston, Texas, United States
[*]Corresponding author. E-mail: kamal.rawal@gmail.com

Abstract. Recent sequencing of genome of northern white-cheeked gibbon (*Nomascus leucogenys*) has provided important insight into fast evolution of gibbons and signatures relevant to gibbon biology. Mobile-genetic elements (MGEs) seem to play a major role in gibbon evolution. Here, we report that most of the gibbon genome is occupied by MGEs such as Alus, MIRs, LINE1, LINE2, LINE3, ERVL, ERV-classI, ERV-classII and other DNA elements which include hAT Charlie and TcMar tigger. We provide detailed description and genome-wide distribution of all the MGEs present in the gibbon genome. Previously, it was reported that gibbon-specific retrotransposon (LAVA) tends to insert into chromosome segregation genes and alter transcription by providing a premature termination site, suggesting a possible molecular mechanism for the genome plasticity of the gibbon lineage. We show that insertion sites of LAVA elements present atypical signals/patterns which are different from typical signals present at the insertion sites of Alu elements. This suggests the possibility of a distinct insertion mechanism used by LAVA elements for their insertions. We also find similarity in the signals of LAVA element insertion sites with atypical signals present at the Alus/L1s insertion sites disrupting the genes leading to diseases such as cancer and Duchenne muscular dystrophy. This suggests the role of LAVA in premature transcription termination.

Keywords. Mobile-genetic elements; LINEs; SINEs; Alus; gibbon; repeat masker; ELAN.

PACS No. 87.18.–h

1. Introduction

Nomascus leucogenys, commonly known as white-cheeked gibbon, is a member of the Hylobatidae family. Gibbons are especially adapted to a tree-living lifestyle with a special form of locomotion called brachiation, or hand-over-hand swinging [1]. Gibbon genome was published in 2014 after sequencing four hominids, which include human, orangutan, gorilla and chimpanzee. They are small arboreal apes closely related to humans and great apes. Comparisons between gibbon chromosomes and those of other primates revealed that gibbons present an extraordinary level of evolutionary chromosomal rearrangements [2]. Mobile-genetic elements (MGEs) account for a significant proportion of eukaryotic genomes and play an important role in altered gene expression and disease. It was shown in a comparative study of gibbons with other primates that genomic breakpoints are significantly depleted of genes and breakpoint intervals contained a mixture of repetitive sequences that are inserted exclusively into the gibbon genome [2]. Carbone *et al.* [2] also reported that the gibbon genome not only contain several classes of transposable elements as described previously in primate genomes but also contain novel retrotransposons called LAVA element [2, 3]. LAVA element is a unique repetitive element that is exclusively present in the gibbon genome and is composed of sequences similar to L1, Alu and VA sections of SVA-mobile elements as shown in figure 1. Earlier studies have also reported that retrotransposons are the most abundant MGEs in mammalian genomes which affect wide-functional activities such as genome evolution, gene disruption and regulation [4, 5]. The L1 family of LINE elements and Alu elements of SINE is the major class of retroelements believed to impact primate genomes such as gorilla, chimpanzee and gibbons [6–8]. These parasitic DNAs have played important roles in the evolution of complex genomes and can also be regarded as useful tools to study organization, structure and function of eukaryotic genomes. They are generally involved in gene inactivation, gene duplication, transduction and regulation [4, 9]. Recent sequencing of gibbon genome has offered an opportunity to study chromosome-wise distribution of these elements. To date, very few studies related to the detailed identification of the transposable elements (TEs) and their insertion sequences in the gibbon genome

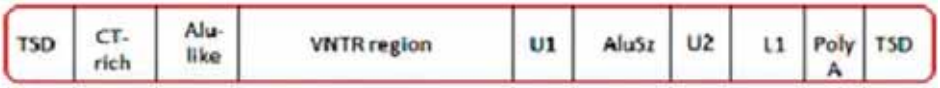

Figure 1. LAVA element structure: LAVA comprises sequences derived from different repetitive elements such as L1, Alu-like and VNTR regions.

have been carried out. The present work involves the identification and classification of different MGEs particularly Alu and LAVA and detailed study of their distribution throughout the genome of *N. leucogenys* using Repeatmasker and ELAN pipeline. ELAN is an efficient computational genomic analysis tool which is responsible for analysing MGEs at different insertion sites whereas Repeatmasker is a program that screens DNA sequences for interspersed repeats [10]. The output of the Repeatmasker program provides detailed classification of the repeats such as LTRs, non-LTRs and other DNA elements. We have also used ELEFINDER (part of ELAN) to find MGEs such as Alu and LAVA in the gibbon genome and also screened the physico-chemical properties of insertion sites using DNA SCANNER.

2. Methods

2.1 *Retrieving genome sequences*

Gibbon genome was retrieved from the NCBI (ftp://ftp.ncbi.nih.gov/genomes/Nomascus_leucogenys/). The chromosomes are made of a total of 2,756,591,777 base pairs of nucleotides.

2.2 *Repeat sequence retrieval*

Identification of TEs based on their features has enabled the construction of libraries of consensus sequences of various types of TEs known as Repbase. Repbase Update (RU) is a comprehensive database of repetitive element consensus sequences [11].

2.3 *Repeatmasker for screening of DNA sequences*

Repeatmasker software screens DNA sequences for low-complexity sequences, repetitive/TEs including small-RNA pseudogenes, Alus, LINEs, SINEs and LTR. Repeatmasker makes use of Repbase libraries which act as the standard reference points for the identification of TEs in a query sequence. It uses statistically optimal scoring matrices derived from the alignments of DNA transposon fossils to their consensus sequences [10]. However, it does not locate all polymorphic simple repetitive sequences. We used RM-BLAST as a search engine in Repeatmasker and tested various options combining speed/sensitivity parameters. Based upon its results, the average length and the total length occupied by each element were also calculated for each chromosome.

2.4 *Whole genome distribution analysis by ELEFINDER*

ELEFINDER finds distribution, nature, orientation, genomic location and site of truncations of MGEs [8, 12]. It not only identifies repeats but also extracts flanking sequence at each MGE site. We used ELAN to find the insertion sites of various MGEs in the gibbon genome. It is a Perl-based system which uses chromosome number, genome file name and element file name as inputs. The output files are generated after performing BLASTN and use of parsers written in Perl. The resulting output files comprise a list of MGEs categorized into full length (intact copies), 5′ truncated, 3′ truncated and both side truncated examples. It also provides upstream sequences for analysis by using a DNA SCANNER tool.

2.5 *DNA SCANNER*

DNA SCANNER tool scans genomic DNA for a variety of properties which include thermodynamic properties, physico-chemical properties, interactions between proteins and sequence-based features such as AT density [8, 13, 14]. It is also responsible for analysing the insertion sites of the repeats that are known and detects the presence of different signals. The DNA sequence analysis method is basically limited to homology searches such as BLAST, finding the patterns in the sequence that are sequence based and the percentages of GC and AT. It offers another pathway for DNA-sequence analysis and has wide pertinence in the analysis of DNA sequences which further allow the analysis of promoters, genic and intergenic regions, insertion sites of retroelements, intron–exon boundaries and clustering sequences based upon their properties [8]. The program evaluates a number of properties for a given DNA sequence in a sliding window to generate sequence-based profiles as explained below.

2.5.1 *Structural signals: DNA bendability:* The ability of a DNA to deform its structural characteristics under a specific stimulus is known as DNA bendability. A tri-nucleotide model based on DNAse-I cutting frequencies predicts that DNAse-I binds and cuts DNA that is bent towards major groove [14, 15].

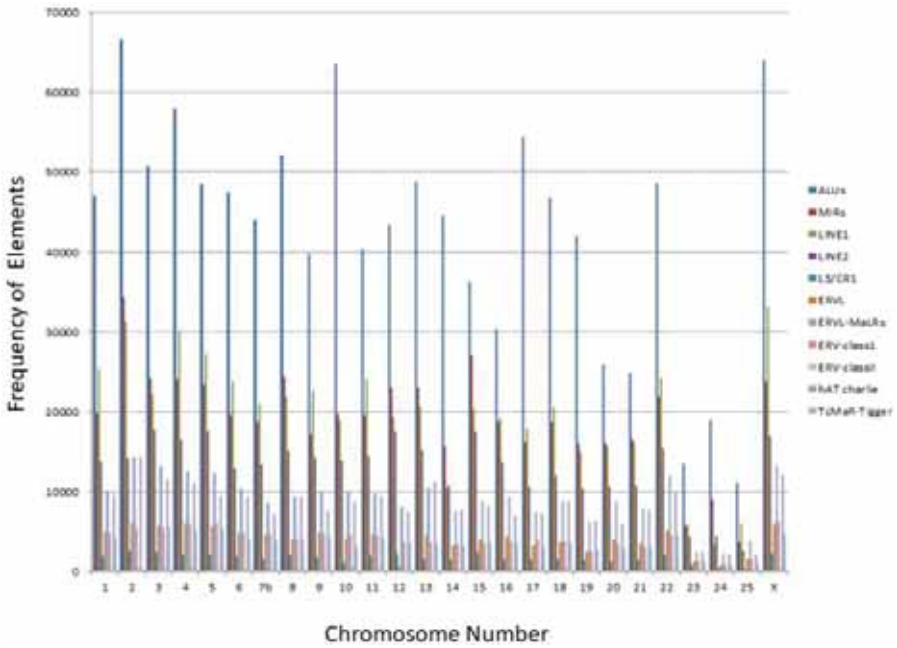

Figure 2. The chromosome-wise representation of the distribution of MGEs in the gibbon genome. Different classes of elements are indicated with different colours. The *y*-axis represents the frequency of elements found on different chromosomes. The *x*-axis represents the chromosome number.

2.5.2 *Thermodynamic signals: Stacking energy:* The stability of a given DNA sequence as well as protein interactions is indicated by the stacking energies which is also responsible for the formation of local structures [14, 16].

2.5.3 *Duplex stability: Free energy signals:* The duplex stability of a DNA structure depends upon its base sequence and more specifically upon 10 different types of nearest-neighbour interactions, namely AA/TT; AT/TA; TA/AT; CA/GT; GT/CA; CT/GA; GA/CT; CG/GC; GC/CG and GG/CC [13, 14].

2.5.4 *Propeller twist signals, bending stiffness and nucleosomal positioning:* DNA must distort in order to bend around a protein: this distortion is facilitated by the deformational capacity of dinucleotides [8]. Some signals are rigid whereas others are flexible, and the propeller twist signals are responsible for measuring the flexibility of DNA [14, 17].

2.5.5 *Protein interaction signals:* The DNA sequences during their interaction with other molecules carry signals specific for their ability to deform [8].

3. Results

Over 43.59% of gibbon genome is occupied by MGEs. The MGEs that were identified in the gibbon genome included the major TE classes that consist of long terminal repeat (LTRs), non-LTRs and DNA transposons. Figure 2 shows the distribution of TEs on different chromosomes of the gibbon genome. Non-LTR transposons were the most abundant TEs in the gibbon genome. Several varieties of superfamilies of LINEs and SINEs were discovered in the gibbon genome. We find that SINEs were present in the form of two superfamilies, namely Alus and MIRs, whereas LINE consisted of LINE1, LINE2 and LINE3/CR1. Among the LINEs, the L1s covered the major portion of the gibbon genome, whereas in the case of SINEs, the Alus were found to be in majority.

LTR retrotransposons were the second most abundant TEs in the gibbon genome and majority of these belong to a class of mammalian repeats derived from retrovirus-like elements. They were categorized into four subgroups, namely ERVL, ERVL-MaLRs, ERV-classI and ERV-classII. DNA transposons are rare in the gibbon genome, and are represented by only two superfamilies (hAT-Charlie and TcMar-Tigger).

Table 1 shows the distribution of the MGEs in different chromosomes of the gibbon genome. The percentage of each chromosome region occupied by repetitive elements was also calculated (table 1). It was found that Alus were the most abundant TEs present in the gibbon, whereas LINE1 was also one of the most frequently occurring MGEs in the genome. Since there are a large number of full-length Alu copies in various genomes, this makes them suitable candidates for detailed analysis as representative SINEs.

Table 1. Distribution of TEs in the gibbon genome.

Chromosome no.	Alus	MIRs	LINE1	LINE2	L3/CR1	ERVL	ERVL-MaLRs	ERV-classI	ERV-classII	TcMaR-	hAT Charlie
1	46974	19825	25315	13750	1823	4803	10160	4703	289	4282	9623
2	66666	34412	31136	14166	2553	5978	14245	5419	328	361	14302
3	50646	24162	22109	17758	2377	5641	13105	5479	335	5574	11543
4	57901	23999	30069	16411	2103	5976	12553	5933	362	5048	11105
5	48480	23364	27250	17580	2113	5686	12323	5953	299	5343	9465
6	47413	19561	23827	12930	1848	4698	10402	4793	265	4050	9323
7b	43972	18707	20885	13407	1503	4520	8521	4606	287	3847	7146
8	52023	24338	21863	15037	1979	3996	9331	4133	260	3933	9281
9	39733	17205	22649	14233	1711	4828	9942	4633	278	4276	7663
10	63538	19727	18953	13869	1078	4068	9890	4479	440	3138	8644
11	40394	19515	24065	14386	1846	4531	9660	4456	247	4196	9485
12	43434	23012	19271	17475	2164	609	8026	3633	271	3624	7457
13	48818	23031	20606	15155	1563	4525	10531	3620	187	3463	11320
14	44633	15638	18468	10696	1401	3264	7591	3501	192	3231	7719
15	36318	27025	20345	17481	2314	3980	8790	3526	187	3640	8082
16	30298	18805	19193	13535	1577	4314	9286	3725	239	3426	6899
17	54354	16149	17856	10632	1371	3308	7339	3890	281	3003	7315
18	46784	18691	20432	12009	1478	3644	8755	3699	226	3547	8747
19	41898	16018	14886	10403	1428	2444	6169	2395	131	2686	6379
20	25857	16027	15686	10641	1170	3948	8805	3349	178	2870	5949
21	24850	16533	16064	10722	1424	3445	7837	3071	170	2973	7768
22	48626	21878	24312	15438	1981	5129	11939	4563	290	4456	10012
23	13509	5688	5892	4236	697	1147	2328	1309	73	1156	2296
24	19032	9032	3408	4457	433	776	2235	824	67	671	2144
25	11113	3568	5878	2618	308	1582	3812	1485	63	917	2130
X	64004	23839	33076	16908	2178	5875	13257	6262	336	4660	12010

4. Analysis of the LINEs and SINEs in gibbon genome

We performed detailed analysis of LINEs, SINEs and LTR elements for each chromosome using Perl scripts and results are summarized in tables 2, 3 and 4. The total number of SINE elements were found to be 16, 07 and 203; LINE elements were 9, 31 and 127 and the total count of DNA elements were found to be 4, 12 and 435.

4.1 *Analysis of LAVA element in gibbon genome*

For each chromosome, the analysis of the LAVA was performed (figures 3–7). It was observed that the signals present in the upstream region were not distinct as observed previously [6, 11, 17, 18]. The insertion sites of LAVA also did not show any characteristic pattern in comparison with signals observed at insertion sites of Alus. The reasons could be attributed to the absence of full-length LAVA elements, the absence of 5′-intact copies or a possibly distinct insertion mechanism in comparison with LINEs or SINEs.

4.2 *Identification of transposable elements using ELAN*

We present insertion sites of Alus in chromosome 24 as a representative case. Signal upstreams of the insertion sites of Alu in chromosome 24 were obtained from DNA SCANNER. It has been observed that Alu elements tend to insert preferentially in the A-rich regions where parameters such as thermodynamic properties, structural features and nucleosomal positioning parameters are different from genomic average values. In contrast, LAVA elements seem to show a different pattern at insertion sites. The peaks or troughs were obtained using DNA SCANNER which compute physicochemical properties using the sliding window mechanism. Different physicochemical properties such as DNA bendability, bending stiffness, protein interaction signals, frequency of AT etc. were computed. The *y*-axis in the profile represents the value of the property and the *x*-axis gives the relative position with respect to the insertion site (figures 8–12).

Table 2. Chromosome-wise summary of the SINE elements in the gibbon genome.

Chromosome no.	Total length	Length occupied	No. of elements	Percentage of sequence (%)
1	124,228,569	15,389,374	67,140	12.39
2	164,513,156	22,689,809	101,439	13.79
3	158,281,636	17,116,535	75,264	10.81
4	154,836,434	18,768,321	82,256	12.12
5	144,113,614	16,384,550	72,266	11.36
6	122,129,856	15,418,127	67,273	12.62
7b	116,018,537	14,380,977	63,042	12.30
8	117,808,995	17,294,411	76,623	14.68
9	118,737,704	13,153,091	57,266	11.07
10	108,686,836	19,369,927	83,444	17.82
11	123,853,985	13,649,407	60,277	11.02
12	109,700,369	14,886,328	66,775	13.50
13	109,844,878	16,187,346	72,089	14.7
14	99,334,225	14,000,626	60,488	14.09
15	110,683,687	13,749,624	63,638	12.42
16	100,086,279	10,888,927	49,331	10.88
17	101,490,674	16,425,099	70,723	16.18
18	105,523,347	9,089,854	65,727	8.61
19	82,726,427	13,371,447	58,100	16.16
20	85,254,692	9,254,553	42,083	10.85
21	87,080,223	9,095,777	41,643	10.44
22	133,238,221	16,175,660	70,911	12.14
23	33,724,709	4,909,473	19,334	14.55
24	28,896,242	6,251,379	28,111	21.63
25	32,264,544	3,484,990	14,759	10.80
X	144,882,752	18,818,075	82,201	12.98

Table 3. Chromosome-wise summary of the LINEs in the gibbon genome.

Chromosome no.	Total length	Length occupied	No. of elements	Percentage of sequence
1	124,228,569	26,587,344	41,596	21.4
2	164,513,156	32,458,804	55,708	19.73
3	158,281,636	33,196,243	51,486	20.97
4	154,836,434	29,872,691	49,351	19.29
5	144,113,614	28,093,514	48,012	19.4
6	122,129,856	23,513,619	38,586	19.25
7b	116,018,537	20,850,749	36,404	17.97
8	117,808,995	21,855,513	39,996	18.5
9	118,737,704	25,798,628	39,250	21.72
10	868,683,610	17,529,436	34,306	16.12
11	123,853,985	26,695,501	41,001	21.55
12	109,700,369	23,387,137	39,489	21.3
13	109,844,878	18,859,033	37,758	17.16
14	99,334,225	17,941,765	31,005	18.06
15	110,683,687	24,119,300	40,681	21.79
16	100,086,279	21,904,114	34,870	21.88
17	101,490,674	17,468,268	30,361	17.21
18	105,523,347	20,511,590	34,469	19.44
19	82,726,427	13,787,352	27,121	16.67
20	85,254,692	16,821,821	27,910	19.73
21	87,080,223	17,954,893	28,634	20.62
22	133,238,221	25,149,791	42,499	18.87
23	33,724,709	6,844,934	11,032	20.2
24	28,896,242	3,087,909	8,407	10.69
25	32,264,544	5,486,671	8,953	17
X	144,882,752	33,881,737	52,842	23.38

Table 4. Chromosome-wise summary of the DNA elements in the gibbon genome.

Chromosome no.	Total length (bp)	Length occupied	No. of elements	Percentage of sequence
1a	124,228,569	4,578,519	18,734	3.69
2	164,513,156	5,981,178	25,705	3.64
3	158,281,636	5,799,716	23,205	3.67
4	154,836,434	5,324,876	21,768	3.43
5	144,113,614	5,116,564	20,062	3.55
6	122,129,856	4,268,588	18,098	3.49
7b	116,018,537	38,518,008	14,800	3.31
8	117,808,995	4,136,569	17,769	3.51
9	118,737,704	4,192,221	16,048	3.53
10	868,683,610	3,441,747	15,483	3.16
11	123,853,985	4,518,862	18,372	3.64
12	109,700,369	3,704,003	15,278	3.37
13	109,844,878	4,361,802	19,462	3.97
14	99,334,225	3,396,239	14,670	3.42
15	110,683,687	3,863,138	15,987	3.49
16	100,086,279	3,567,688	14,056	3.56
17	101,490,674	3,106,632	13,797	3.06
18	105,523,347	3,789,619	16,231	3.59
19	82,726,427	2,818,295	12,191	3.41
20	85,254,692	2,978,639	12,066	3.40
21	87,080,223	3,307,062	14,011	3.79
22	133,238,221	4,821,418	19,763	3.62
23	33,724,709	1,158,301	4,730	3.43
24	28,896,242	751,431	3,739	2.60
25	32,264,544	1,025,009	4,208	3.17
X	144,882,752	5,092,937	22,202	3.51

5. Discussion

Recent sequencing of gibbon has contributed significantly to the understanding the evolution of human, chimpanzee and gorilla. The gibbon genome contains all previously described classes of TEs that are present in other primates as well. The mystery surrounding TEs revolves around their mode and selection of insertion sites. Different categories of TEs use different mechanisms to get inserted into the genome which include target prime retrotransposition. Previously, gibbon–great ape population divergence time was estimated to be of ~16.8 million years ago assuming a split time with a macaque of 29 million years ago; therefore, the LAVA element is believed to be originated around the time of the divergence of gibbons from the ancestral great ape or human lineage. The

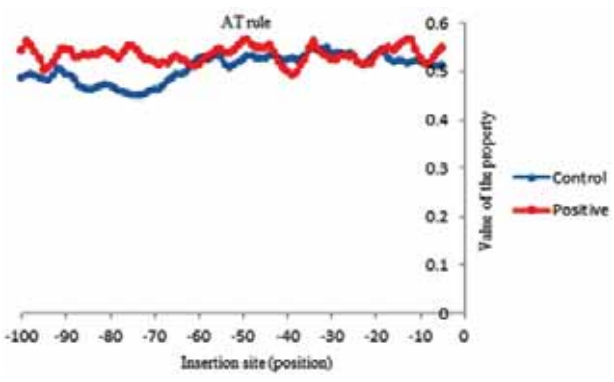

Figure 3. AT frequency profile for the upstream region of LAVA insertions in chromosome 24.

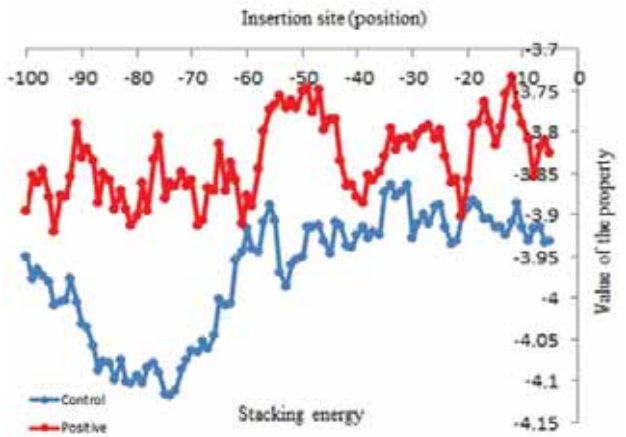

Figure 4. Stacking energy profile for upstream region of LAVA insertions in chromosome 24.

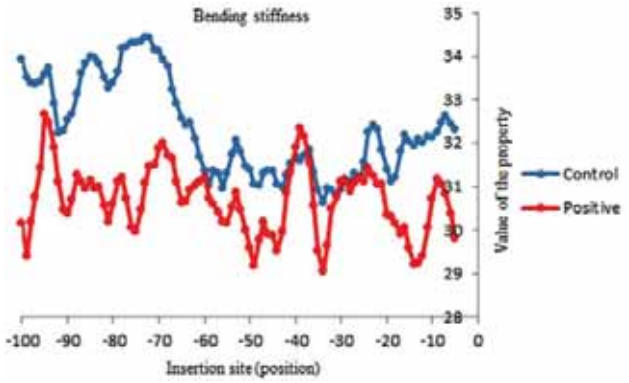

Figure 5. Bending stiffness profile for the upstream region of LAVA insertions in chromosome 24.

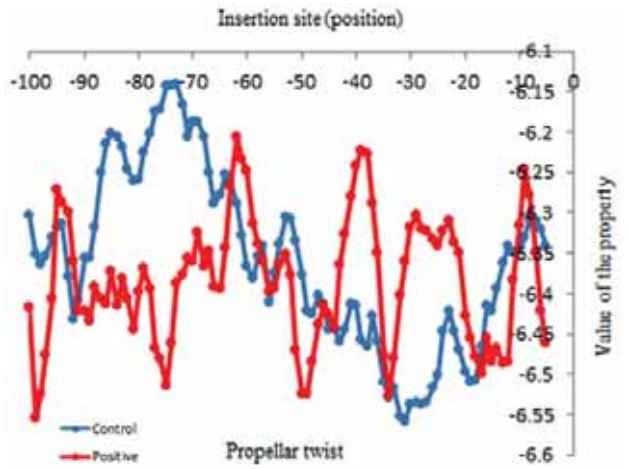

Figure 6. Propeller twist profile for the upstream region of LAVA insertions in chromosome 24.

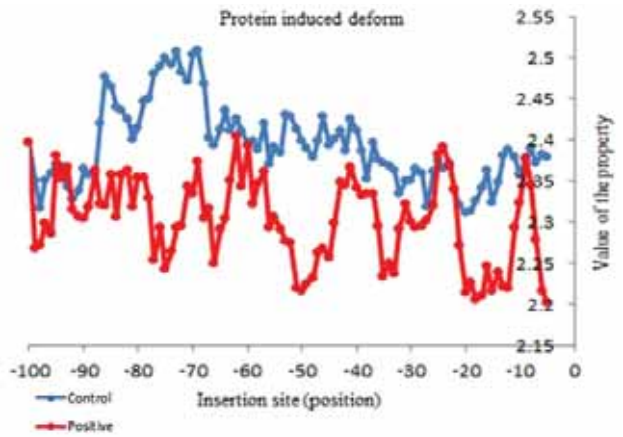

Figure 7. Protein-induced deformability profile for the upstream region of LAVA insertions in chromosome 24.

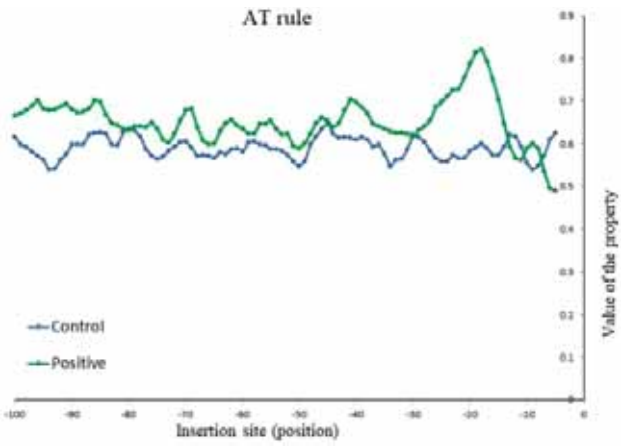

Figure 8. Adenine thymine density profile for the insertion sites present upstream of full-length Alu insertions on 24th chromosome of gibbon computed by DNA SCANNER. The *y*-axis represents the value of the property under study (referred to as 'AT Rule') and the *x*-axis represents the position with respect to the insertion site (taken as position 0). Here, blue curve shows the profile for the control dataset (randomly selected genomic sites) and green curve shows the profile for the positive dataset.

LAVA element is implicated for the mechanism to cause gross-chromosomal changes leading to accelerated evolution. We have used Repeatmasker and ELAN to find different MGEs from the gibbon genome. In ELAN, ELEFINDER and DNA SCANNER were used for detailed analysis of insertions of Alu and LAVA elements in the genome and to scan the insertion sites of gibbon chromosomes for different physicochemical properties for protein interactions [8]. We observed that a significant difference in patterns or signals is seen in the insertion sites of Alu and LAVA elements. This suggests that LAVA element tends to insert in a mechanism which is distinct from the mechanism and property of LINEs and SINEs. Previously, we have shown that signals observed in the insertion sites of Alu elements are different when inserted in the exonic region, intronic as well as intergenic regions [12]. The signals observed in the intergenic regions are conserved across various Alu insertion sites in various primate genomes [12, 14, 19]. Previously, we conducted detailed analyses of genes in research which included genes such as APC, Duchenne myotrophy disease (DMD) and CYBB for Alu, L1 and SVA insertions that were reported to be responsive to TE insertion leading to disease. In our study, we found that the signals observed in most of the disease genes were atypical in nature like in DMD which is one of the largest genes known (2.4 Mb in length). Also, it was seen that there is an occurrence of L1 insertion which has been reported to cause disruption in the gene which further leads to DMD.

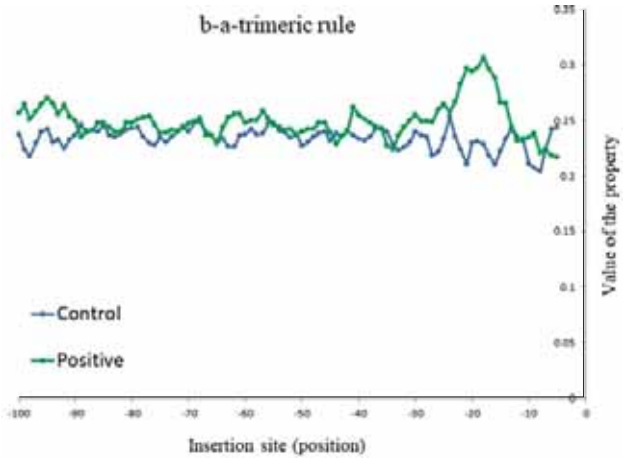

Figure 9. B to A form density parameter profile for the insertion sites present upstream of full-length Alu insertions in chromosome 24 of gibbon genome.

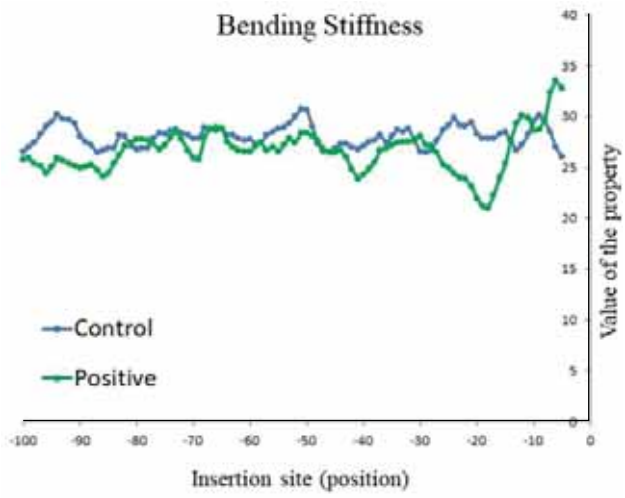

Figure 10. Bending stiffness profile for the insertion sites present upstream of full-length Alu insertions in chromosome 24 of gibbon genome.

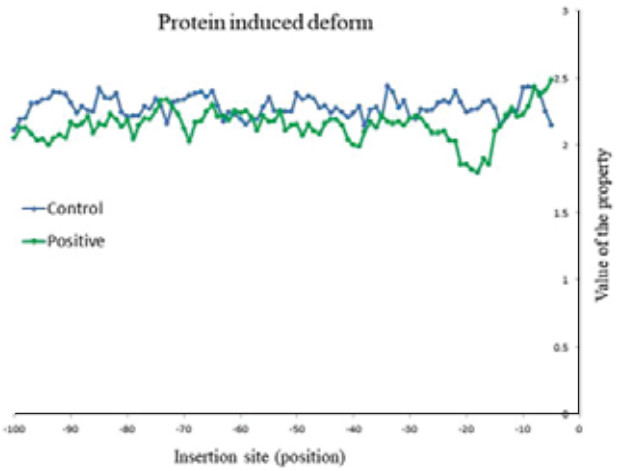

Figure 11. Protein-induced deformability profile generated for the insertion sites present upstream of full-length Alu insertions in chromosome 24 of gibbon genome.

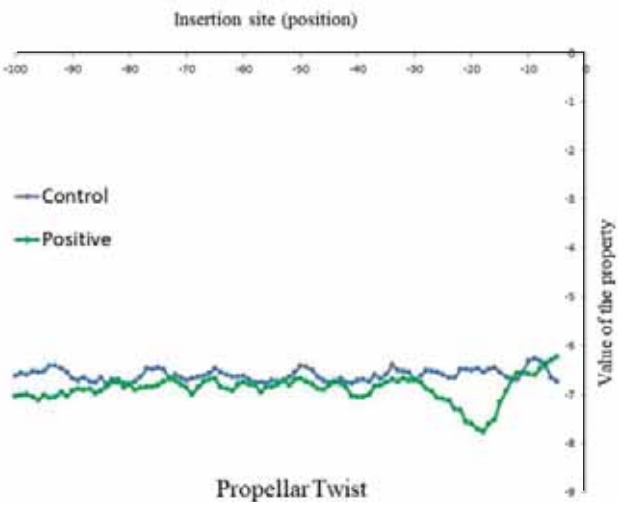

Figure 12. Propeller twist profile for the insertion sites present upstream of full-length Alu insertions in chromosome 24 of gibbon genome.

This and other examples can be found online at *http://nldsps.jnu.ac.in/TE-Vs-diseases/database/*. As per our study of 49 genes, we assume that the absence of typical signals within the exonic regions prevents further wide-spread disruption of coding regions by Alu and L1s respectively. Similarly, it was suggested by [2] that the high level of chromosomal rearrangements in the gibbon genome may have been due to the premature transcription termination mechanism of the segregated genes of the chromosomes due to LAVA activities. Though this mechanism occurred at a very low level but still it was sufficient to increase the chromosome segregation errors [9]. Hence, it is interesting to observe that the possible genetic target sites of LAVA element present atypical signals that are quite similar to the signals observed in human genes amenable to disease causing MGE insertions. Importantly, these signals were conserved in genomes of the lower eukaryotes such as *Entamoeba histolytica* [19, 20].

Acknowledgements

The authors acknowledge the contribution of their lab members during the preparation of this manuscript: Hitesh Kumar Jaiswal, Samridhi Mehta, Garima Chaudhary, Mitika Raj, Srashti Varshney and Shubhani

Chokra. The authors thank the Department of Biotechnology, Government of India, for the financial support (Grant Id: BT/PR17252/BID/7/708/2016). They also thank the Robert J Kleberg, Jr. and Helen C Kleberg Foundation, USA and Baylor College of Medicine, Houston USA for grant support.

References

[1] R Roberto *et al.*, *Genome Res.* **17**, 249 (2007)
[2] L Carbone *et al.*, *Nature* **513**, 195 (2014)
[3] B McClintock, *Cold Spring Harbor Symposia on Quantitative Biology* **21**, 197 (1956)
[4] C Bergman and H Quesneville, *Briefings Bioinfm.* **8**, 382 (2007)
[5] R Cordaux and M Batzer, *Nat. Rev. Genet.* **10**, 691 (2009)
[6] D Finnegan, *Curr. Biol.* **7**, R245 (1997)
[7] J Moran, R J DeBerardinis and H H Kazazian, *Science* **283(5407)**, 1530 (1999)
[8] K Rawal and R Ramaswamy, *Nucl. Acids Res.* **39**, 6864 (2011)
[9] H H Kazazian, *Nat. Genet*. **22**, 130 (1999)
[10] A F A Smit and P Green, http://repeatmasker.org
[11] J Jurka, V Kapitonov, A Pavlicek, P Klonowski, O Kohany and J Walichiewicz, *Research*, **110**, 462 (2005)
[12] K Rawal, A Priya, A Malik, R Bahl and R Ramaswamy, *Mobile Genet. Elem.* **2**, 133 (2012).
[13] P Mandal, K Rawal, R Ramaswamy, A Bhattacharya and S Bhattacharya, *Nucl. Acids Res.* **34**, 5752 (2006)
[14] K Rawal, S Dorji, A Kumar, A Ganguly and A Grewal, *Mobile Genet. Elem.* **3**, e25675 (2013)
[15] D M Crothers, T E Haran and J G Nadeau, *J. Biol. Chem.* **265**, 7093 (1990)
[16] S G Delcourt and R D Blake, *J. Biol. Chem.* **266**, 15160 (1991)
[17] E M Hassan and C Calladine, *J. Mol. Biol.* **259**, 95 (1996)
[18] R Harshey and A Bukhari, *Proceedings of the National Academy of Sciences* **78**, 1090 (1981)
[19] A Bakre, K Rawal, R Ramaswamy, A Bhattacharya and S Bhattacharya, *Exp. Parasitol.* **110**, 207 (2005)
[20] B Dev, A Malik and K Rawal, *Bioinformation* **8**, 777 (2012)