



MAPPING AND INTERPRETATION OF REPETITIVE INGREDIENTS IN *VITIS VINIFERA* (GRAPES).

Virendra Kumar*, Mohd. Faheem Khan, Lalit Kumar Pandey and A.K. Jaitly

Department of Plant Science, Mahatma Jyotiba Phule Rohilkhand University, Bareilly (U.P.) India.

Abstract

Repetitive sequences are ubiquitous segments of plants genome. The occurrence of repetitive sequences within the noncoding or coding region of DNA probably encouraging changes in the overall structure of a functional genome that makes them an emerging and very influential object for a broad range of molecular variation in plants which would be potentially severe for fitness and phenotypic consequences. Last decayed research shows the adverse impact of repetitive elements raises on genome organization due to harsh environmental exposure. Therefore, for the better judgment of transposable elements role, the magnitude of their dispersion in the grape genome by the computational approach has been analyzed. The analysis revealed more than forty percent genome of *Vitis vinifera* comprises a total 3,79,530 copies of different superfamilies transposable elements (TEs). The TEs copies also distinguished according to their status in plant genome whether they are truncated or intact on both ends. The 680 copies of Copia and 933 copies of Harbinger superfamily elements are completely inserted (intact on both ends) and the remaining element copies partially inserted (i.e. intact on one and truncated on other end or truncated on both ends) in *Vitis vinifera* genome. Out of 379530 copies of transposable elements, 13290 copies present in gene sequences and however 9042 copies present in promoter regions of different genes.

Key words: Repetitive elements, Dispersion, Simple Sequence Repeats, Transposable elements, Copies.

Introduction

The majority of the plants comprise a large fraction of repetitive sequences and its correlation with organism genome size represents a key discovery of genomics biology (Cruz *et al.*, 2016; Negi *et al.*, 2016). During environmental stimuli the repetitive sequences shows an enormous impact on evolution and stability of genome or regulation of gene expression due to implication in numerous processes such as chromosome movement, recombination, rearrangement, interaction of chromatin and histone proteins with the determination of chromosome (Mehrotra and Goyal, 2014; Macas *et al.*, 2015; Cruz *et al.*, 2016). Development of Next-generation sequencing technology and genome sequencing efforts in the plant kingdom has produced approximately 55 plus genomes sequencing information to date (Michael and Jackson, 2013), which provide novel insights for the investigation of a wide range of biological questions into genome organization and function (Cheng *et al.*, 2016). The research of last decades shows the amplification of repeats can influence gene expression and function

(Dodsworth *et al.*, 2015; Macas *et al.*, 2015).

Transposable elements (TEs) were named after their characteristics that they transpose, change locations in a genome and increase copy numbers (Naito *et al.*, 2014). TEs have been divided into two main classes according to their method of transposition and each class divided into subclasses according to the different molecules used as intermediates in the replicative mechanism of transposition and the nature of the transposing unit (Negi *et al.*, 2016). Class I (Retrotransposons) elements transpose *via* an RNA intermediate by reverse transcription and subsequent integration while Class II elements (DNA Transposons) transpose *via* a DNA intermediate and possess terminal inverted repeats (Piégue *et al.*, 2015). There are two major subclasses of class I elements: (1) Long terminal repeat (LTR) retrotransposons, include gypsy-like and Copia-like two major superfamilies; (2) non-Long terminal repeat (non-LTR) retrotransposons, which include long and short interspersed elements (LINEs and SINEs). Class II elements (DNA transposons) are divided into two subclasses (Wicker *et al.*, 2007). Subclass I include TEs

*Author for correspondence : E-mail: virenmtech@gmail.com

Table 1: Number of the Copies of different family elements present in *Vitis vinifera* genome along with their genome coverage.

Class	Subclass	Order	Superfamily	Families	Copies	Genome Coverage
Retrotransposon		LTR	Gypsy	48	260934	30.012
			Copia	72	26635	3.59
			Caulimoviridae	1	1107	0.215
		Non-LTR	LINE/L1	6	9846	1.13
Total				127	298522	34.947
DNA Transposon	Subclass - I	TIR	Harbinger	10	8888	0.41
			hAT	10	7695	0.667
			Mutator/MuDR	14	48632	2.879
			EnSpm/CACTA	10	9478	1.148
	Subclass -II	Helitron	Helitron	1	40	0.003
Unclassified DNA transposon			DNA Transposon	1	3679	0.147
Total				46	78412	5.254
Unclassified TE		Transposable Element		1	2596	0.1

that transpose by an excision and repair (cut and paste) method using a transposase that recognizes their terminal inverted repeat (TIR) sequences. Plant TEs that belong to superfamilies, hAT, MuDR, Harbinger and EnSpm/CACTA are part of this subclass. Helitrons belong to subclass II, replicated by rolling circle mechanism. Furthermore, Harbinger gave rise to miniature inverted-repeat transposable elements (MITEs).

The ultimate role of TEs in plant genomes remains controversial, however, it is clear that they are involved in regulatory networks/mechanism and contribute to genome evolution (Cruz *et al.*, 2016). Knowledge of the distribution, genomic organization, chromosomal location and evolutionary origin of repetitive DNA sequences is necessary for insight into the organization, evolution, behavior and functional potential (Mehrotra and Goyal, 2014). The ultimate role of repetitive elements in plant genomes remains controversial. Indeed, they were proven to be a crucial tool for gene expression in many organisms and also an effective way to develop molecular markers. *Vitis vinifera*, a member of Vitaceae family, is a widely cultivated species around the world including India and a good system for functional genomics of fruiting plants. The crop has wide adaptability and can be grown under temperate, subtropical and tropical climatic and varied agro-ecological settings. The ~500 Mbp long genome of *Vitis vinifera* organized into 19 chromosomes. In this study some parts of chromosome represented as Chr_1r, 3r, 4r, 5r, 7r, 9r, 10r, 11r, 12r, 13r, 16r, 17r, 18r which is the random sequence of the respective chromosome. However, approximately 43 Mbp remains uncharacterized (Jaillon *et al.*, 2007).

Therefore the *Vitis vinifera* genome was analyzed in order to (1) examine the distribution of TEs along with their status about the completeness of the element; and (2) the copies of TEs present in genes or promoter regions. The main aim of the current study is to characterize repeats dispersion and assemble a comprehensive dataset of inserted transposable elements with the status of their 5' and 3' ends, either they are intact or truncated and their genomic coverage.

Materials and Methods

The complete genome of *Vitis vinifera* available in public domain National Centre of Biotechnology Information NCBI (<http://www.ncbi.nlm.nih.gov>) was used for this study. The pre-identified repeats of *Vitis vinifera* were collected from Repbase (Bao *et al.*, 2015). The research work was performed in four segments. In the first segment, the generic feature format (GFF3) file was processed by Bedtools and Perl script to know the coverage of exon, intron and intergenic portions of the genome. In the second segment in house bash scripts

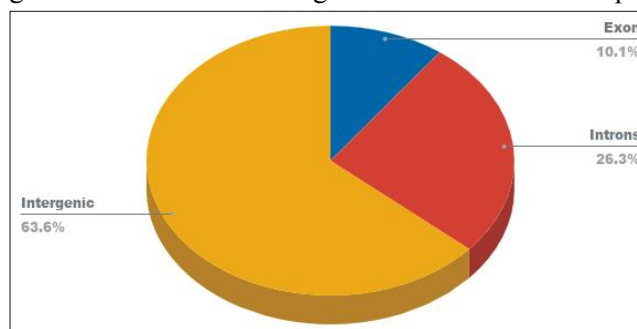


Fig. 1: Genomic Portion of *Vitis vinifera* covered by exon, introns and intergenic segments.

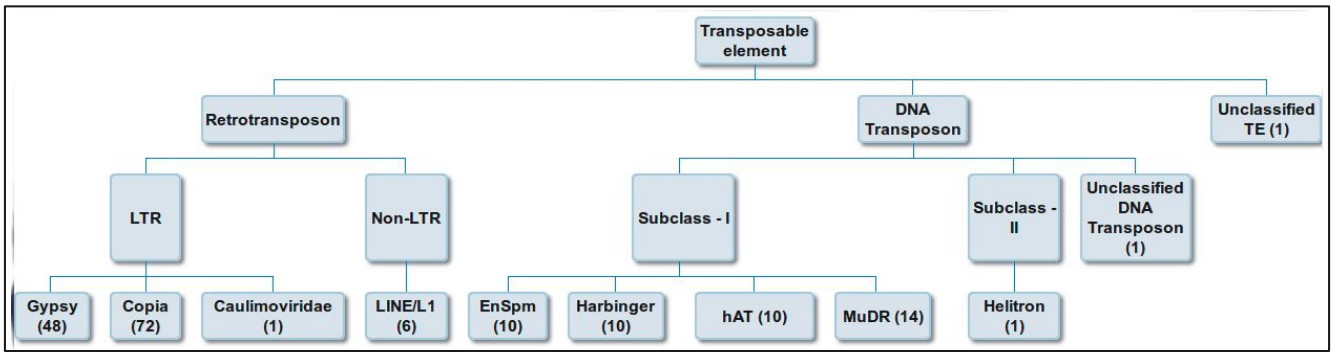


Fig. 2: Classification of Transposable elements of *Vitis vinifera*.

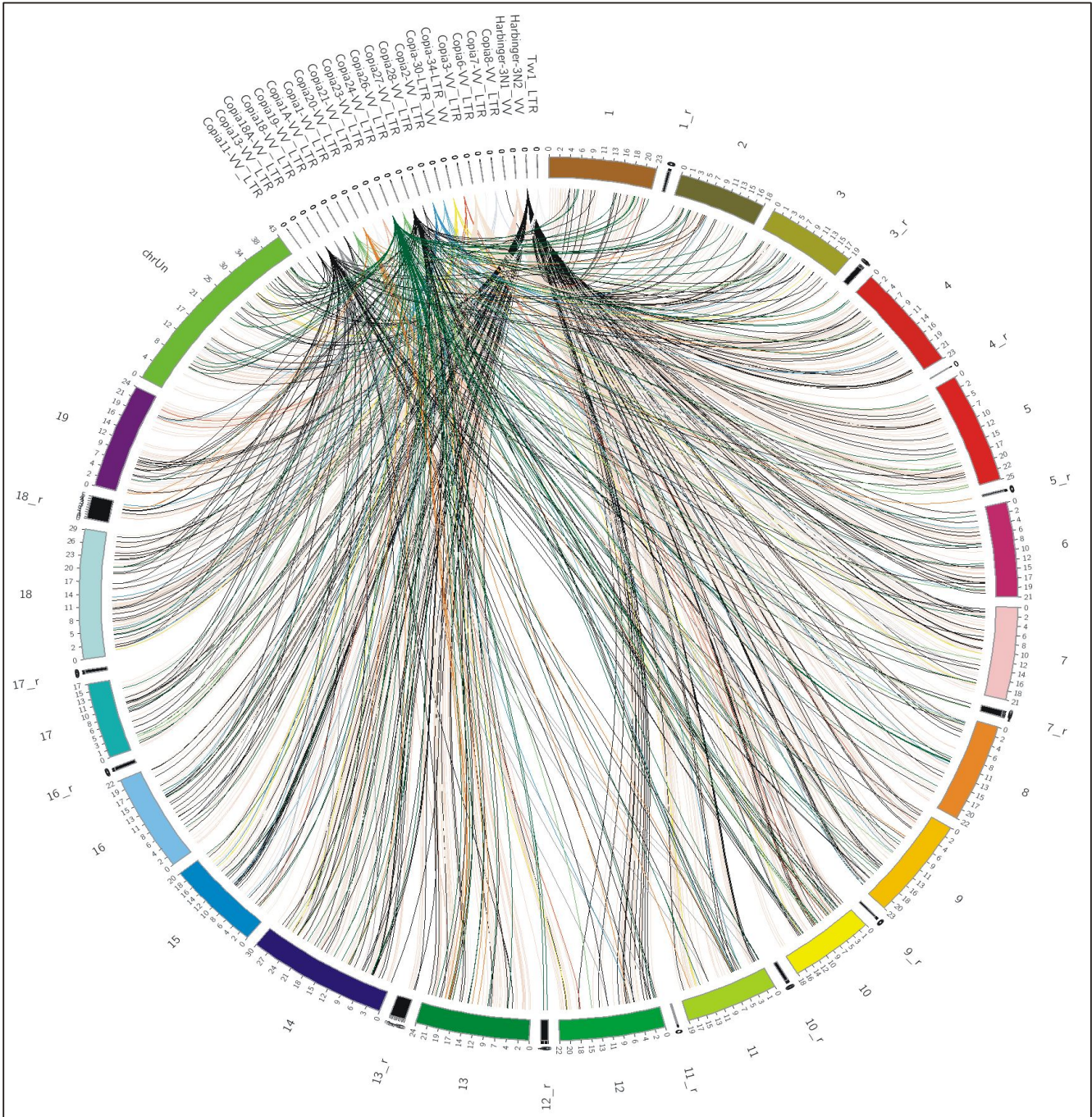


Fig. 3: Distribution of completely inserted elements copies of different Superfamilies on 19 chromosome of *Vitis vinifera*.

Table 2: Number of different superfamily TEs present on each chromosome of *Vitis vinifera*.

Chr. No.	Element Superfamily										
	Copia	Gypsy	Caulim- oviridae	L1	EnSpm	hAT	Harb- inger	MuDR	Heli- tron	Unclssified DNA Transposon	Unclassified TE
Chr_1	1200	12113	52	435	454	367	454	3262	2	161	117
chr_1r	45	1503	-	1	8	-	-	7	-	1	2
Chr_2	853	11610	43	363	407	253	358	1393	7	130	98
Chr_3	1038	13676	48	262	451	299	377	2606	-	157	134
chr_3r	88	774	3	43	13	39	12	86	-	3	15
Chr_4	1382	9897	107	553	383	319	508	2359	4	187	124
chr_4r	16	1	-	-	10	1	2	11	-	-	4
Chr_5	1296	13519	37	731	434	447	456	3451	3	176	176
chr_5r	49	29	2	36	2	26	5	17	-	2	4
Chr_6	956	7776	31	505	462	317	450	2260	1	180	141
Chr_7	1174	8153	43	426	445	241	476	1278	5	171	72
chr_7r	94	508	3	22	6	39	20	133	-	9	7
Chr_8	1070	8274	83	480	266	296	347	2637	-	204	114
Chr_9	1278	9954	50	316	415	456	438	2652	-	250	110
chr_9r	14	885	-	-	48	10	4	45	-	-	6
Chr_10	1241	11320	49	364	373	302	303	1416	-	115	70
chr_10r	43	263	-	25	3	-	9	10	-	4	2
Chr_11	1225	12809	21	606	281	281	355	1408	-	162	108
chr_11r	25	6	-	13	3	2	1	4	-	-	5
Chr_12	1225	13330	53	433	417	382	393	1615	-	189	114
chr_12r	129	1163	13	56	9	25	21	202	-	3	9
Chr_13	1335	13493	80	589	349	451	412	1395	-	195	90
chr_13r	175	1462	8	95	122	47	49	551	-	39	14
Chr_14	1813	15030	44	753	605	499	586	3810	5	196	124
Chr_15	1370	8893	18	350	386	334	384	1936	-	158	109
Chr_16	1179	13640	46	310	528	502	474	1777	2	207	151
chr_16r	30	473	-	6	1	7	8	3	-	-	2
Chr_17	937	5366	36	434	316	266	284	1378	-	112	78
chr_17r	44	1136	2	3	1	4	4	22	-	5	1
Chr_18	1580	16300	67	529	471	462	586	1902	4	177	181
chr_18r	294	4206	-	54	278	137	90	907	-	21	41
Chr_19	1526	14263	39	339	548	368	437	2960	4	227	158
chr_Un	1911	29109	129	714	983	516	585	5139	3	238	215

Table 3: Number of TEs copies status in *Vitis vinifera* genome.

Gypsy						
S. No.	Element Name	OK_OK	OK_Trunc	Trunc_OK	Trunc_Trunc	Total
1	Gret1_I	NA	91	NA	1257	1348
2	Gret1_LTR	NA	380	NA	762	1142
3	Gypsy1-VV_I	NA	10	NA	170	180
4	Gypsy1-VV_LTR	NA	8	NA	21	29
5	Gypsy10-VV_I	NA	5	NA	31	36
6	Gypsy10-VV_LTR	NA	4	NA	NA	4
7	Gypsy11-VV_I	NA	486	NA	9145	9631
8	Gypsy11-VV_LTR	NA	732	NA	5907	6639
9	Gypsy12-VV_I	NA	269	NA	8340	8609

Table 3 Continue ...

were developed which uses stand-alone BLAST to track the genome-wide dispersion of previously characterized repetitive elements (Tao, 2016). The third segment was performed to track the genome-wide dispersion of previously characterized TEs in *Vitis vinifera* with the help of the script developed in the second segment of this work. The bash script also processed the BLAST output by a series of filters to assemble the significant information under the captions like as serial number, element name, chromosome number, element

Continue Table 3 ...

10	Gypsy12-VV_LTR	NA	4257	NA	5456	9713
11	Gypsy13-VV_I	NA	489	NA	8312	8801
12	Gypsy13-VV_LTR	NA	732	NA	6598	7330
13	Gypsy14-VV_I	NA	417	NA	7520	7937
14	Gypsy14-VV_LTR	NA	5082	NA	9556	14638
15	Gypsy15-VV_I	NA	9	NA	11	20
16	Gypsy15-VV_LTR	NA	17	NA	12	29
17	Gypsy16-VV_I	NA	81	NA	1159	1240
18	Gypsy16-VV_LTR	NA	370	NA	791	1161
19	Gypsy17-VV_I	NA	134	NA	7541	7675
20	Gypsy17-VV_LTR	NA	5257	NA	10128	15385
21	Gypsy18-VV_I	NA	51	NA	320	371
22	Gypsy18-VV_LTR	NA	178	NA	721	899
23	Gypsy19-VV_I	NA	627	NA	8791	9418
24	Gypsy19-VV_LTR	NA	5250	NA	13712	18962
25	Gypsy2-VV_I	NA	45	NA	88	133
26	Gypsy2-VV_LTR	NA	112	NA	454	566
27	Gypsy20-VV_I	NA	394	NA	6770	7164
28	Gypsy20-VV_LTR	NA	5196	NA	9254	14450
29	Gypsy22-VV_I	NA	177	NA	1022	1199
30	Gypsy22-VV_LTR	NA	5	NA	1772	1777
31	Gypsy3-VV_I	NA	180	NA	1344	1524
32	Gypsy3-VV_LTR	NA	305	NA	1480	1785
33	Gypsy4-VV_I	NA	745	NA	7367	8112
34	Gypsy4-VV_LTR	NA	5237	NA	8898	14135
35	Gypsy5-VV_I	NA	6	NA	6	12
36	Gypsy5-VV_LTR	NA	14	NA	1	15
37	Gypsy6-VV_I	NA	123	NA	9886	10009
38	Gypsy6-VV_LTR	NA	5053	NA	10949	16002
39	Gypsy7-VV_I	NA	742	NA	8264	9006
40	Gypsy7-VV_LTR	NA	5294	NA	11622	16916
41	Gypsy8-VV_I	NA	6	NA	135	141
42	Gypsy8-VV_LTR	NA	13	NA	NA	13
43	Gypsy9-VV_I	NA	148	NA	337	485
44	Gypsy9-VV_LTR	NA	512	NA	749	1261
45	GYVIT1_I	NA	307	NA	7475	7782
46	GYVIT1_LTR	NA	4447	NA	10463	14910
47	VI_I	NA	233	NA	164	397
48	VI_LTR	NA	520	NA	1423	1943
Copia						
1	Copia-29-I_VV	NA	3	NA	149	152
2	Copia-29-LTR_VV	NA	28	NA	82	110
3	Copia-30-I_VV	NA	10	NA	5	15
4	Copia-30-LTR_VV	20	3	2	NA	25
5	Copia-31-I_VV	NA	200	NA	142	342
6	Copia-31-LTR_VV	NA	404	NA	146	550
7	Copia-32-I_VV	NA	34	NA	44	78
8	Copia-33-I_VV	NA	272	NA	662	934
9	Copia-33-LTR_VV	NA	405	NA	774	1179
10	Copia-34-I_VV	NA	41	NA	91	132

Table 3 Continue ...

start, element end, strand (plus/minus), chromosome length, target start, target end and further description about the completeness of the element. In the fourth segment, the repeat locations compared with Gene transfer format (GTF) file of *Vitis vinifera* using BEDTools, to obtain the copies of transposable elements present in gene and promoter regions (Aaron *et al.*, 2010). The Libre Office and Circos tool was used to the graphical representation of results (Krzywinski *et al.*, 2009).

Results and Discussion

The analysis of *Vitis vinifera* GFF3 file revealed only 26.26 percent genome covered by introns however 10.13 percent covered by exons and a remaining large fraction (63.61%) of the plant genome is Intergenic (Fig. 1).

The computational analysis of *Vitis vinifera* genome revealed the ununiform distribution of 3,79,530 copies of TEs, which cover approximately 40 percent of plant genome. The TEs copies classified according to their class, subclass, order, superfamily, family (Table 1). There are 174 repetitive elements were identified in the *Vitis vinifera* genome. Of these, 72 belonging to the Copia superfamily, 10 belonging to EnSpm, 48 belonging to Gypsy, Harbinger and hAT both has 10 elements, 6 belonging to L1, 14 belonging to MuDR, Helitron and Caulimoviridae each has one element. There are two elements which are not fit in this classification one of them belong to DNA transposon and another one is unclassified Transposable Element (Fig. 2).

The analysis of *Vitis vinifera* genome revealed the Gypsy as a most abundant (68.75%) superfamily followed by MuDR (12.81%), Copia (7.02%), L1 (2.59%), EnSpm (2.50%), Harbinger (2.34%), hAT (2.03%), DNA transposon (0.97%), TES (0.68%), Caulimoviridae subfamily (0.29%) and 40 copies were identified

Continue Table 3 ...

11	Copia-34-LTR_VV	9	9	38	89	145
12	Copia-35-I_VV	NA	7	NA	284	291
13	Copia-35-LTR_VV	NA	19	NA	NA	19
14	Copia1-VV_I	NA	31	NA	375	406
15	Copia1-VV_LTR	30	36	2	21	89
16	Copia10-VV_I	NA	240	NA	1589	1829
17	Copia10-VV_LTR	NA	1277	NA	1594	2871
18	Copia11-VV_I	NA	320	NA	1627	1947
19	Copia11-VV_LTR	2	755	110	415	1282
20	Copia12-VV_I	NA	45	NA	165	210
21	Copia12-VV_LTR	NA	134	NA	5	139
22	Copia13-VV_I	NA	8	NA	31	39
23	Copia13-VV_LTR	9	4	2	3	18
24	Copia15-VV_I	NA	26	NA	169	195
25	Copia15-VV_LTR	NA	71	NA	17	88
26	Copia16-VV_I	NA	3	NA	377	380
27	Copia16-VV_LTR	NA	17	NA	13	30
28	Copia17-VV_I	NA	9	NA	651	660
29	Copia18-VV_I	NA	73	NA	274	347
30	Copia18-VV_LTR	52	66	103	52	273
31	Copia18A-VV_I	NA	65	NA	507	572
32	Copia18A-VV_LTR	68	63	85	52	268
33	Copia19-VV_I	NA	13	NA	16	29
34	Copia19-VV_LTR	12	14	NA	3	29
35	Copia1A-VV_I	NA	20	NA	204	224
36	Copia1A-VV_LTR	12	20	1	12	45
37	Copia2-VV_I	NA	11	NA	617	628
38	Copia2-VV_LTR	6	20	10	19	55
39	Copia20-VV_I	NA	11	NA	21	32
40	Copia20-VV_LTR	14	14	6	1	35
41	Copia21-VV_I	NA	4	NA	49	53
42	Copia21-VV_LTR	7	NA	NA	1	8
43	Copia22-VV_I	NA	63	NA	317	380
44	Copia22-VV_LTR	NA	74	NA	182	256
45	Copia23-VV_I	NA	30	NA	1291	1321
46	Copia23-VV_LTR	136	685	2	29	852
47	Copia24-VV_I	NA	9	NA	44	53
48	Copia24-VV_LTR	20	11	NA	NA	31
49	Copia26-VV_I	NA	87	NA	2246	2333
50	Copia26-VV_LTR	72	144	27	83	326
51	Copia27-VV_I	NA	11	NA	575	586
52	Copia27-VV_LTR	6	18	4	9	37
53	Copia28-VV_I	NA	25	NA	109	134
54	Copia28-VV_LTR	19	34	2	6	61
55	Copia29-VV_I	NA	11	NA	315	326
56	Copia29-VV_LTR	NA	24	NA	18	42
57	Copia3-VV_I	NA	174	NA	626	800
58	Copia3-VV_LTR	150	21	158	46	375
59	Copia4-VV_I	NA	24	NA	388	412
60	Copia4-VV_LTR	NA	6	NA	2	8

Table 3 Continue ...

as members of Helitron superfamily (Table 1). The analysis revealed Class I (Retrotransposon) element with 298522 copies covers 34.95% genome of *Vitis vinifera*. However, with 78412 copies Class II (DNA transposon) elements covered 5.25% of the *Vitis vinifera* genome, 2596 copies belongs to unclassified transposable elements and covers 0.1% of total genome (Table 1). Out of class I (Retrotransposon) elements, Gypsy superfamilies elements are most abundant (260934) followed by Copia (26635), L1/LINE (9846) and Caulimoviridae (1107) elements. However, in case of class II elements, Mutator/MuDR detected as a most abundant (48632) followed by EnSpm (9478), Harbinger (8888), hAT (7695), Unclassified DNA transposon (3679) and Helitron (40). The number of each superfamily copies *i.e.* Caulimoviridae, Copia, DNA Transposon, EnSpm, Gypsy, Harbinger, hAT, Helitron, L1, MuDR and Other Transposable Element on each chromosome given in table 2.

The resultant data processed for completeness of elements in each superfamily and assemble into tabulated form. The highest populated Gypsy type of element was observed as Gypsy19-VV_LTR which has 18962 copies distributed throughout the genome (Table 3). The following order represents the population density of each and every Gypsy type of retroelements scattered in the grapes genome *i.e.* Gypsy7-VV_LTR > Gypsy6-VV_LTR > Gypsy17-VV_LTR > GYVIT1_LTR > Gypsy14-VV_LTR > Gypsy20-VV_LTR > Gypsy4-VV_LTR > Gypsy6-VV_I > Gypsy12-VV_LTR > Gypsy11-VV_I > Gypsy19-VV_I > Gypsy7-VV_I > Gypsy13-VV_I > Gypsy12-VV_I > Gypsy4-VV_I > Gypsy14-VV_I > GYVIT1_I > Gypsy17-VV_I > Gypsy13-VV_LTR > Gypsy20-VV_I > Gypsy11-VV_LTR > V1_LTR > Gypsy3-

Continue Table 3 ...

61	Copia5-VV_I	NA	39	NA	83	122
62	Copia5-VV_LTR	NA	58	NA	51	109
63	Copia6-VV_I	NA	13	NA	131	144
64	Copia6-VV_LTR	1	24	3	5	33
65	Copia7-VV_I	NA	10	NA	62	72
66	Copia7-VV_LTR	7	2	13	2	24
67	Copia8-VV_I	NA	7	NA	24	31
68	Copia8-VV_LTR	1	13	NA	2	16
69	Copia9-VV_I	NA	20	NA	219	239
70	Copia9-VV_LTR	NA	30	NA	225	255
71	Tvv1_I	NA	24	NA	371	395
72	Tvv1_LTR	27	45	15	22	109
Caulimoviridae						
1	CAULIV1	NA	206	NA	901	1107
LINE/L1						
1	VLINE1_VV	NA	NA	NA	1088	1088
2	VLINE2_VV	NA	32	NA	2173	2205
3	VLINE3_VV	NA	2	NA	2167	2169
4	VLINE4_VV	NA	7	NA	586	593
5	VLINE5_VV	NA	2	NA	2934	2936
6	VLINE6_VV	NA	22	NA	833	855
Harbinger						
1	Harbinger-1_VV	NA	76	NA	659	735
2	Harbinger-1N1_VV	NA	84	NA	399	483
3	Harbinger-3_VV	NA	13	NA	590	603
4	Harbinger-3N1_VV	638	359	287	714	1998
5	Harbinger-3N2_VV	295	551	241	692	1779
6	Harbinger-3N3_VV	NA	8	NA	45	53
7	Harbinger-3N3B_VV	NA	5	NA	12	17
8	VHARB-N2_VV	NA	293	NA	292	585
9	VHARB-N3_VV	NA	22	NA	1802	1824
10	VHARB4_VV	NA	164	NA	647	811
hAT						
1	hAT-10_VV	NA	2	NA	691	693
2	hAT-10N1_VV	NA	120	NA	299	419
3	hAT-11N_VV	NA	3	NA	354	357
4	hAT-6_VV	NA	22	NA	205	227
5	hAT-7_VV	NA	22	NA	344	366
6	VIHAT1	NA	27	NA	2867	2894
7	VIHAT2	NA	93	NA	817	910
8	VIHAT2-N1_VV	NA	103	NA	607	710
9	VIHAT3	NA	96	NA	465	561
10	VIHAT3-N1_VV	NA	104	NA	454	558
MuDR						
1	MuDR-11N_VV	NA	7	NA	333	340
2	MuDR-12_VV	NA	74	NA	1546	1620
3	MuDR-13_VV	NA	96	NA	23475	23571
4	MuDR-18_VV	NA	31	NA	2412	2443
5	MuDR-21_VV	NA	70	NA	4841	4911
6	MuDR-3_VV	NA	84	NA	1194	1278

Table 3 Continue ...

VV_LTR > Gypsy22-VV_LTR > Gypsy3-VV_I > Gret1_I > Gypsy9-VV_LTR > Gypsy16-VV_I > Gypsy22-VV_I > Gypsy16-VV_LTR > Gret1_LTR > Gypsy18-VV_LTR > Gypsy2-VV_LTR > Gypsy9-VV_I > V1_I > Gypsy18-VV_I > Gypsy1-VV_I > Gypsy8-VV_I > Gypsy2-VV_I > Gypsy10-VV_I > Gypsy1-VV_LTR > Gypsy15-VV_LTR > Gypsy15-VV_I > Gypsy5-VV_LTR > Gypsy8-VV_LTR > Gypsy5-VV_I and Gypsy10-VV_LTR has minimum population. Similarly the highest populated Copia type of element was observed as Copia10-VV_LTR which has 2871 copies distributed throughout in the genome as shown in table 3. The following decreasing order represents the number of inserted copies observed for Copia superfamily elements *i.e.* Copia10-VV_LTR > Copia26-VV_I > Copia11-VV_I > Copia10-VV_I > Copia23-VV_I > Copia11-VV_LTR > Copia-33-LTR_VV > Copia-33-I_VV > Copia23-VV_LTR > Copia3-VV_I > Copia17-VV_I > Copia2-VV_I > Copia27-VV_I > Copia18A-VV_I > Copia-31-LTR_VV > Copia4-VV_I > Copia1-VV_I > Tvv1_I > Copia16-VV_I > Copia22-VV_I > Copia3-VV_LTR > Copia18-VV_I > Copia-31-I_VV > Copia26-VV_LTR > Copia29-VV_I > Copia-35-I_VV > Copia18-VV_LTR > Copia18A-VV_LTR > Copia22-VV_LTR > Copia9-VV_LTR > Copia9-VV_I > Copia1A-VV_I > Copia12-VV_I > Copia15-VV_I > Copia-29-I_VV > Copia-34-LTR_VV > Copia6-VV_I > Copia12-VV_LTR > Copia28-VV_I > Copia-34-I_VV > Copia5-VV_I > Copia-29-LTR_VV > Copia5-VV_LTR > Tvv1_LTR > Copia1-VV_LTR > Copia15-VV_LTR > Copia-32-I_VV > Copia7-VV_I > Copia28-VV_LTR > Copia2-VV_LTR > Copia21-VV_I > Copia24-VV_I > Copia1A-VV_LTR > Copia29-VV_LTR > Copia13-VV_I > Copia27-VV_LTR > Copia20-VV_LTR > Copia6-VV_LTR > Copia20-VV_I > Copia24-VV_LTR > Copia8-VV_I > Copia16-VV_LTR > Copia19-VV_I >

Continue Table 3 ...

7	MuDR-4_VV	NA	46	NA	1267	1313
8	MuDR-5_VV	NA	1	NA	494	495
9	MuDR-6_VV	NA	50	NA	3293	3343
10	MuDR-7_VV	NA	14	NA	32	46
11	MuDR-8_VV	NA	57	NA	1289	1346
12	MuDR-9_VV	NA	104	NA	797	901
13	MUDRAV11	NA	187	NA	5400	5587
14	MUDRAV12	NA	43	NA	1395	1438
EnSpm						
1	EnSpm-13_VV	NA	1	NA	1035	1036
2	EnSpm-3_VV	NA	17	NA	902	919
3	EnSpm-4_VV	NA	245	NA	1292	1537
4	EnSpm-4N1_VV	NA	245	NA	972	1217
5	EnSpm-5_VV	NA	109	NA	2394	2503
6	EnSpm-6_VV	NA	47	NA	529	576
7	EnSpm-8N_VV	NA	9	NA	35	44
8	ENSPM-N3_VV	NA	48	NA	168	216
9	ENSPM1_VV	NA	56	NA	443	499
10	ENSPM2_VV	NA	34	NA	897	931
Helitron						
1	Helitron1_VV	NA	2	NA	38	40
Unclassified DNA Transposon						
1	VHARB-N1_VV	NA	917	NA	2762	3679
Unclassified Transposable Element						
1	TE-7-1_VV	NA	117	NA	2479	2596

Table 4: Number of TEs copies detected in genes of *Vitis vinifera* and their status.

Gypsy						
S. No.	Element Name	OK OK	OK Trun	Trun OK	Trun Trun	Total
1	Gret1_I	NA	2	NA	32	34
2	Gret1_LTR	NA	11	NA	36	47
3	Gypsy1-VV_I	NA	3	NA	15	18
4	Gypsy1-VV_LTR	NA	2	NA	NA	2
5	Gypsy10-VV_I	NA	1	NA	1	2
6	Gypsy11-VV_I	NA	2	NA	36	38
7	Gypsy11-VV_LTR	NA	8	NA	96	104
8	Gypsy12-VV_I	NA	1	NA	34	35
9	Gypsy12-VV_LTR	NA	57	NA	87	144
10	Gypsy13-VV_I	NA	2	NA	38	40
11	Gypsy13-VV_LTR	NA	8	NA	123	131
12	Gypsy14-VV_I	NA	3	NA	41	44
13	Gypsy14-VV_LTR	NA	69	NA	157	226
14	Gypsy16-VV_I	NA	3	NA	19	22
15	Gypsy16-VV_LTR	NA	12	NA	37	49
16	Gypsy17-VV_I	NA	NA	NA	48	48
17	Gypsy17-VV_LTR	NA	71	NA	154	225
18	Gypsy18-VV_I	NA	NA	NA	5	5
19	Gypsy18-VV_LTR	NA	7	NA	45	52
20	Gypsy19-VV_I	NA	2	NA	49	51

Table 4 Continue ...

Copia19-VV LTR > Copia-30-LTR VV > Copia7-VV LTR > Copia-35-LTR VV > Copia13-VV LTR > Copia8-VV LTR > Copia-30-I VV > Copia21-VV LTR and Copia4-VV LTR. Out of 26635 copies of copia element, 680 copies are both ends intact, 6541 copies 5 prime intact and 3 prime truncated, 583 copies 5 prime truncated and 3 prime intact and 18831 copies are truncated on both ends. The Copia3-VV_LTR is the element of Copia superfamily which shows highest dispersion of completely inserted element.

Caulimoviridae is a superfamily of sudotransposable elements; they are molecular fossils of Endogenous viral elements. In-silico analysis of Caulimoviridae elements in *Vitis vinifera* genome shows colonization. Total 1107 copies dispersed throughout the *Vitis vinifera* genome and cover approximately 0.215% of the genome. Out of 1107 copies, 901 copies are truncated at both ends however the remaining 206 copies intact at 5 prime locations and truncated at 3 prime locations.

Long interspersed nuclear elements / LINE-1 (L1) superfamily elements belong to the long interspersed non-Long terminal repeat retrotransposons. The genome-wide analysis of these elements dispersion revealed a total of 9846 copies. Out of which 9781 copies truncated on both ends and remaining 65 copies insert on 5 prime and truncated at 3 prime ends. The number of LINE-1 elements *Vitis vinifera* genome is in following decreasing order i.e., VLINE5 VV > VLINE2 VV > VLINE3 VV > VLINE1 VV > VLINE6 VV > VLINE4 VV. The result shows that non-LTR retrotransposon elements VLINE5 VV has maximum copies inserted and VLINE4 VV has the lowest number of copies inserted throughout in *Vitis vinifera* genome.

Continue Table 4 ...

21	Gypsy19-VV_LTR	NA	71	NA	183	254
22	Gypsy2-VV_I	NA	1	NA	9	10
23	Gypsy2-VV_LTR	NA	5	NA	34	39
24	Gypsy20-VV_I	NA	2	NA	41	43
25	Gypsy20-VV_LTR	NA	70	NA	148	218
26	Gypsy22-VV_I	NA	4	NA	19	23
27	Gypsy22-VV_LTR	NA	NA	NA	57	57
28	Gypsy3-VV_I	NA	5	NA	25	30
29	Gypsy3-VV_LTR	NA	11	NA	50	61
30	Gypsy4-VV_I	NA	4	NA	43	47
31	Gypsy4-VV_LTR	NA	71	NA	139	210
32	Gypsy6-VV_I	NA	NA	NA	62	62
33	Gypsy6-VV_LTR	NA	70	NA	167	237
34	Gypsy7-VV_I	NA	2	NA	47	49
35	Gypsy7-VV_LTR	NA	71	NA	167	238
36	Gypsy8-VV_I	NA	2	NA	10	12
37	Gypsy8-VV_LTR	NA	3	NA	1	4
38	Gypsy9-VV_I	NA	7	NA	23	30
39	Gypsy9-VV_LTR	NA	48	NA	76	124
40	GYVIT1_I	NA	2	NA	37	39
41	GYVIT1_LTR	NA	65	NA	156	221
42	VI_I	NA	10	NA	13	23
43	VI_LTR	NA	49	NA	146	195
Copia						
1	Copia-29-I_VV	NA	NA	NA	5	5
2	Copia-29-LTR_VV	NA	NA	NA	9	9
3	Copia-31-I_VV	NA	11	NA	5	16
4	Copia-31-LTR_VV	NA	23	NA	5	28
5	Copia-32-I_VV	NA	1	NA	3	4
6	Copia-33-I_VV	NA	8	NA	30	38
7	Copia-33-LTR_VV	NA	8	NA	30	38
8	Copia-34-I_VV	NA	NA	NA	4	4
9	Copia-34-LTR_VV	NA	NA	1	2	3
10	Copia-35-I_VV	NA	3	NA	95	98
11	Copia-35-LTR_VV	NA	5	NA	NA	5
12	Copia1-VV_I	NA	10	NA	93	103
13	Copia1-VV_LTR	11	11	NA	4	26
14	Copia10-VV_I	NA	2	NA	15	17
15	Copia10-VV_LTR	NA	35	NA	43	78
16	Copia11-VV_I	NA	93	NA	454	547
17	Copia11-VV_LTR	NA	213	24	119	356
18	Copia12-VV_I	NA	7	NA	15	22
19	Copia12-VV_LTR	NA	18	NA	1	19
20	Copia13-VV_I	NA	NA	NA	3	3
21	Copia15-VV_I	NA	NA	NA	6	6
22	Copia15-VV_LTR	NA	1	NA	1	2
23	Copia16-VV_I	NA	NA	NA	14	14
24	Copia17-VV_I	NA	NA	NA	18	18
25	Copia18-VV_I	NA	21	NA	70	91
26	Copia18-VV_LTR	19	24	33	19	95

Table 4 Continue ...

In-silico analysis of class II elements (DNA transposons) in *Vitis vinifera* revealed that 78412 copies of DNA transposons randomly dispersed and comprises approximately 5.25 percent of the total plant genome. The class II elements (DNA transposons) of *Vitis vinifera* further classified into two subclasses i.e. Subclass I and Subclass II. There are four superfamilies which contain terminal inverted repeats (TIR) i.e. EnSpm, MuDR, Harbinger, hAT belongs to Subclass I of DNA Transposon of *Vitis vinifera*, however, Subclass II contains only superfamily Helitron. There is a DNA Transposon which does not fit in this classification.

The EnSpm is a superfamily of autonomous DNA transposons of subclass I of class II. A total of 9478 copies of EnSpm were observed throughout the genome, of which 8867 copies truncated on both ends and the remaining 811 copies intact on 5 prime and truncated at 3 prime (Table 3). Family wise number of EnSpm transposon element copies was detected in following decreasing order i.e., EnSpm-5 VV > EnSpm-4 VV > EnSpm-4N1 VV > EnSpm-13 VV > ENSPM2 VV > EnSpm-3 VV > EnSpm-6 VV > ENSPM1 VV > ENSPM-N3 VV > EnSpm-8N VV which shows EnSpm-5VV element has maximum copies inserted and EnSpm-8N VV has the lowest number of copies inserted throughout the genome. Similarly the decreasing order represents the number of inserted copies observed for Harbinger superfamily is following Harbinger-3N1 VV > VHARB-N3 VV > Harbinger-3N2 VV > VHARB4 VV > Harbinger-1 VV > Harbinger-3 VV > VHARB-N2 VV > Harbinger-1N1 VV > Harbinger-3N3 VV > Harbinger-3N3B VV. Results show the dispersion of 8888 copies of Harbinger superfamily out of which 933 copies inserted on both ends, 1575 copies 5 prime intact and 3

Continue Table 4 ...

27	Copia18A-VV_I	NA	20	NA	99	119
28	Copia18A-VV_LTR	28	22	25	11	86
29	Copia19-VV_I	NA	NA	NA	3	3
30	Copia19-VV_LTR	1	NA	NA	1	2
31	Copia1A-VV_I	NA	5	NA	54	59
32	Copia1A-VV_LTR	5	6	NA	6	17
33	Copia2-VV_I	NA	5	NA	106	111
34	Copia2-VV_LTR	2	10	4	13	29
35	Copia20-VV_I	NA	NA	NA	1	1
36	Copia22-VV_I	NA	4	NA	46	50
37	Copia22-VV_LTR	NA	1	NA	12	13
38	Copia23-VV_I	NA	12	NA	344	356
39	Copia23-VV_LTR	25	210	NA	8	243
40	Copia24-VV_I	NA	1	NA	1	2
41	Copia24-VV_LTR	3	NA	NA	NA	3
42	Copia26-VV_I	NA	NA	NA	50	50
43	Copia26-VV_LTR	NA	3	1	3	7
44	Copia27-VV_I	NA	2	NA	154	156
45	Copia27-VV_LTR	1	3	1	1	6
46	Copia28-VV_I	NA	1	NA	15	16
47	Copia28-VV_LTR	1	1	NA	1	3
48	Copia29-VV_I	NA	5	NA	58	63
49	Copia29-VV_LTR	NA	7	NA	4	11
50	Copia3-VV_I	NA	45	NA	166	211
51	Copia3-VV_LTR	54	5	35	11	105
52	Copia4-VV_I	NA	3	NA	102	105
53	Copia5-VV_I	NA	NA	NA	1	1
54	Copia5-VV_LTR	NA	NA	NA	1	1
55	Copia6-VV_I	NA	1	NA	5	6
56	Copia6-VV_LTR	NA	1	NA	NA	1
57	Copia7-VV_I	NA	1	NA	3	4
58	Copia7-VV_LTR	1	NA	1	NA	2
59	Copia9-VV_I	NA	1	NA	3	4
60	Copia9-VV_LTR	NA	2	NA	8	10
61	Tvv1_I	NA	3	NA	88	91
62	Tvv1_LTR	4	9	3	6	22
Caulimoviridae						
1	CAULIV1	NA	4	NA	37	41
LINE/L1						
1	VLINE1_VV	NA	NA	NA	354	354
2	VLINE2_VV	NA	11	NA	639	650
3	VLINE3_VV	NA	1	NA	513	514
4	VLINE4_VV	NA	2	NA	185	187
5	VLINE5_VV	NA	1	NA	929	930
6	VLINE6_VV	NA	8	NA	247	255
Harbinger						
1	Harbinger-1_VV	NA	4	NA	29	33
2	Harbinger-1N1_VV	NA	4	NA	17	21
3	Harbinger-3_VV	NA	1	NA	61	62
4	Harbinger-3N1_VV	71	23	25	42	161

Table 4 Continue ...

prime truncated, 528 copies intact at 5 prime and truncated at 3 prime, the remaining 5852 copies truncated on both ends. The highest populated Harbinger type of element *i.e.* Harbinger-3N1 VV has 1998 copies distributed throughout in the genome out of these 638 copies are intact on both ends.

Members of hAT (half-a-tetratricopeptide) superfamily belong to Class II transposable elements and move by excision and integration. These elements copies number was detected in following decreasing order *i.e.*, VIHAT1 > VIHAT2 > VIHAT2-N1 VV > hAT-10 VV > VIHAT3 > VIHAT3-N1 VV > hAT-10N1 VV > hAT-7 VV > hAT-11N VV > hAT-6 VV. Status of each family element of hAT superfamily also given into table 3. A total of 7695 randomly distributed copies of hAT superfamily in grapes genome were observed. Out of 7695 copies, 7103 truncated on both ends and the remaining 592 copies intact at 5 prime and truncated at 3 prime.

MuDR superfamily also belongs to non-autonomous class II transposable elements. Status of each family of MuDR superfamily was detected and shows the following decreasing order on the bases of element copies *i.e.*, MuDR-13 VV > MUDRAV11 > MuDR-21 VV > MuDR-6 VV > MuDR-18 VV > MuDR-12 VV > MUDRAV12 > MuDR-8 VV > MuDR-4 VV > MuDR-3 VV > MuDR-9 VV > MuDR-5 VV > MuDR-11N VV > MuDR-7 VV. The analysis revealed a total of 48632 copies randomly distributed throughout the genome. Out of which 47768 copies truncated on both ends and the remaining 864 copies intact at 5 prime and truncated at 3 prime locations.

Helitrons superfamily belongs to the class of non-autonomous DNA transposon and uses rolling-circle transposition mechanisms for

Continue Table 4 ...

5	Harbinger-3N2_VV	36	44	17	49	146
6	Harbinger-3N3_VV	NA	2	NA	13	15
7	Harbinger-3N3B_VV	NA	3	NA	4	7
8	VHARB-N2_VV	NA	27	NA	9	36
9	VHARB-N3_VV	NA	4	NA	165	169
10	VHARB4_VV	NA	14	NA	41	55
hAT						
1	hAT-10_VV	NA	NA	NA	33	33
2	hAT-10N1_VV	NA	16	NA	25	41
3	hAT-11N_VV	NA	NA	NA	32	32
4	hAT-6_VV	NA	NA	NA	6	6
5	hAT-7_VV	NA	2	NA	61	63
6	VIHAT1	NA	NA	NA	106	106
7	VIHAT2-N1_VV	NA	16	NA	65	81
8	VIHAT3-N1_VV	NA	30	NA	105	135
MuDR						
1	MuDR-11N_VV	NA	NA	NA	2	2
2	MuDR-12_VV	NA	NA	NA	63	63
3	MuDR-13_VV	NA	6	NA	266	272
4	MuDR-18_VV	NA	2	NA	139	141
5	MuDR-21_VV	NA	1	NA	135	136
6	MuDR-3_VV	NA	2	NA	30	32
7	MuDR-4_VV	NA	1	NA	37	38
8	MuDR-5_VV	NA	NA	NA	19	19
9	MuDR-6_VV	NA	NA	NA	143	143
10	MuDR-8_VV	NA	3	NA	58	61
11	MuDR-9_VV	NA	NA	NA	31	31
12	MUDRAV11	NA	2	NA	194	196
13	MUDRAV12	NA	2	NA	43	45
EnSpm						
1	EnSpm-13_VV	NA	NA	NA	17	17
2	EnSpm-3_VV	NA	1	NA	15	16
3	EnSpm-4_VV	NA	9	NA	38	47
4	EnSpm-4N1_VV	NA	9	NA	46	55
5	EnSpm-5_VV	NA	3	NA	113	116
6	EnSpm-6_VV	NA	NA	NA	16	16
7	ENSPM-N3_VV	NA	1	NA	6	7
8	ENSPM1_VV	NA	2	NA	25	27
9	ENSPM2_VV	NA	1	NA	14	15
Unclassified DNA Transposon						
1	VHARB-N1_VV	NA	111	NA	200	311
Unclassified Transposable Element						
1	TE-7-1_VV	NA	19	NA	175	194

transposition. The In-silico analysis estimated 40 copies of Helitron1 VV element in *Vitis vinifera* genome, which covers approximately 0.003 megabases of the host genome. Status of these family elements given into table 3. Out of 40, 38 copies truncated on both ends and remaining two copies intact at 5' prime and truncated at 3' prime.

There was a DNA transposon namely VHARB-N1 VV of *Vitis vinifera* not follow the above-given classification, therefore, placed separately as an unclassified DNA Transposon. The genome-wide analysis revealed the dispersion of 3679 copies of VHARB-N1 VV element in the plant genome. Further analysis of these elements about their status in the plant genome revealed out of 3679 copies, 2762 copies truncated on both ends and the remaining 917 copies intact at 5' prime and truncated at 3' prime locations (Table 3).

The family-wise number of elements copies and their status is given into table 3. However, the completely inserted elements copies (680 of Copia and 933 of Harbinger superfamily) on chromosomes of *Vitis vinifera* graphically represented with the help of Circos tool in fig. 3.

Repeats Present in gene sequences

The GTF file of *Vitis vinifera* was processed to obtain TEs copies in genes and revealed there is only 9.86 % (*i.e.* 2623) genes comprise 13290 copies of TEs however no any insertion in 23965 genes. The 1010 genes out of 2623 show a single copy of transposable elements however remaining 1613 genes shows two or more than two copies of Transposable elements. TEs copies present in gene sequences analyzed for superfamily wise number and status of element copies either they are intact on both ends, truncated on both ends, intact at 5' prime end or intact at 3' prime. The analyses revealed Copia superfamily as a most abundant with 3614 copies followed by Gypsy (3543), L1 (2890), MuDR (1179), Harbinger (705), hAT (497), EnSpm (316) unclassified DNA transposon (311), Unclassified transposable elements (191), Caulimoviridae (41) superfamily in *Vitis vinifera* gene (Table 4). In the tabulated data 262 copies (*i.e.* 155 copies of Copia and 105 copies of harbinger) are intact on both ends (Completely inserted), 2112 copies

Table 5: Number TEs copies and their status in promoters of *Vitis vinifera* genes.

Gypsy						
S. No.	Element Name	OK_OK	OK_Trun	Trun_OK	Trun_Trun	Total
1	Gret1_I	NA	1	NA	44	45
2	Gret1_LTR	NA	33	NA	79	112
3	Gypsy1-VV_I	NA	NA	NA	3	3
4	Gypsy1-VV_LTR	NA	NA	NA	1	1
5	Gypsy11-VV_I	NA	1	NA	32	33
6	Gypsy11-VV_LTR	NA	5	NA	84	89
7	Gypsy12-VV_I	NA	NA	NA	26	26
8	Gypsy12-VV_LTR	NA	46	NA	77	123
9	Gypsy13-VV_I	NA	1	NA	29	30
10	Gypsy13-VV_LTR	NA	5	NA	101	106
11	Gypsy14-VV_I	NA	NA	NA	26	26
12	Gypsy14-VV_LTR	NA	53	NA	148	201
13	Gypsy15-VV_I	NA	NA	NA	1	1
14	Gypsy15-VV_LTR	NA	1	NA	1	2
15	Gypsy16-VV_I	NA	2	NA	31	33
16	Gypsy16-VV_LTR	NA	30	NA	79	109
17	Gypsy17-VV_I	NA	NA	NA	28	28
18	Gypsy17-VV_LTR	NA	57	NA	142	199
19	Gypsy18-VV_I	NA	NA	NA	7	7
20	Gypsy18-VV_LTR	NA	14	NA	33	47
21	Gypsy19-VV_I	NA	1	NA	27	28
22	Gypsy19-VV_LTR	NA	57	NA	162	219
23	Gypsy2-VV_I	NA	NA	NA	7	7
24	Gypsy2-VV_LTR	NA	9	NA	44	53
25	Gypsy20-VV_I	NA	1	NA	23	24
26	Gypsy20-VV_LTR	NA	54	NA	122	176
27	Gypsy22-VV_I	NA	3	NA	41	44
28	Gypsy22-VV_LTR	NA	NA	NA	109	109
29	Gypsy3-VV_I	NA	3	NA	39	42
30	Gypsy3-VV_LTR	NA	28	NA	128	156
31	Gypsy4-VV_I	NA	2	NA	26	28
32	Gypsy4-VV_LTR	NA	55	NA	119	174
33	Gypsy6-VV_I	NA	NA	NA	31	31
34	Gypsy6-VV_LTR	NA	53	NA	165	218
35	Gypsy7-VV_I	NA	2	NA	26	28
36	Gypsy7-VV_LTR	NA	58	NA	160	218
37	Gypsy8-VV_I	NA	NA	NA	1	1
38	Gypsy8-VV_LTR	NA	1	NA	NA	1
39	Gypsy9-VV_I	NA	4	NA	17	21
40	Gypsy9-VV_LTR	NA	21	NA	35	56
41	GYVIT1_I	NA	NA	NA	30	30
42	GYVIT1_LTR	NA	48	NA	145	193
43	VI_I	NA	4	NA	9	13
44	VI_LTR	NA	22	NA	48	70
Copia						
1	Copia-29-I_VV	NA	NA	NA	5	5
2	Copia-29-LTR_VV	NA	NA	NA	22	22
3	Copia1-VV_I	NA	1	NA	26	27
4	Copia1-VV_LTR	NA	2	NA	3	5

Table 5 Continue ...

5' end intact and 3' end truncated, 170 copies 5' end truncated and 3' end intact and 10746 copies truncated at both ends.

Repeats Present in promoter regions

The analysis of promoter regions (3K upstream and 2K downstream) for dispersion of repeats revealed the presence of 9042 copies of TEs in 2925 gene promoter. The 1478 promoters show a single copy of TEs however remaining 1447 promoters comprises two or more than two transposable elements. The transposable elements copies of promoter regions also characterized according to their status either they are completely intact, truncated on both ends, intact at 5' prime end or intact at 3' prime end. The analysis revealed Gypsy as a most most abundant superfamily with 3161 copies in promoters followed by MuDR (2131), Harbinger (1279), Copia (906), EnSpm (405), hAT (353), DNA transposon (301), L1 (299), TEs (185), Caulimoviridae (21) and one copy of Helitron superfamily (Table 5). In the tabulated data 139 copies (*i.e.* 26 copies of Copia and 113 copies of harbinger) are both ends intact, 1294 copies 5' end intact and 3' end truncated, 90 copies 5' end truncated and 3' end intact and 7519 copies truncated at both ends.

Conclusion

Repetitive elements are ancient, widespread, diverse, abundant and probably independently evolved from their hosts. The TEs propagate through a sophisticated interaction with the ecosystem and often guided by opposite selective forces, used to mitigating their deleterious effects on the host genome. In this study, the availability of the *Vitis vinifera* genome sequence has allowed the scanning of the whole genome for the distribution of TEs. Also highlights the distribution of TEs

Continue Table 5 ...

5	Copia10-VV_I	NA	NA	NA	47	47
6	Copia10-VV_LTR	NA	39	NA	85	124
7	Copia11-VV_I	NA	3	NA	29	32
8	Copia11-VV_LTR	NA	18	4	14	36
9	Copia12-VV_I	NA	4	NA	15	19
10	Copia12-VV_LTR	NA	11	NA	1	12
11	Copia13-VV_I	NA	NA	NA	1	1
12	Copia15-VV_I	NA	NA	NA	13	13
13	Copia15-VV_LTR	NA	4	NA	4	8
14	Copia16-VV_I	NA	NA	NA	22	22
15	Copia16-VV_LTR	NA	2	NA	2	4
16	Copia17-VV_I	NA	NA	NA	25	25
17	Copia18-VV_I	NA	1	NA	12	13
18	Copia18-VV_LTR	3	2	4	2	11
19	Copia18A-VV_I	NA	1	NA	22	23
20	Copia18A-VV_LTR	3	3	4	1	11
21	Copia1A-VV_I	NA	NA	NA	10	10
22	Copia1A-VV_LTR	1	NA	1	NA	2
23	Copia20-VV_I	NA	NA	NA	2	2
24	Copia20-VV_LTR	2	NA	NA	NA	2
25	Copia22-VV_I	NA	7	NA	14	21
26	Copia22-VV_LTR	NA	4	NA	7	11
27	Copia23-VV_I	NA	NA	NA	36	36
28	Copia23-VV_LTR	3	15	NA	2	20
29	Copia24-VV_I	NA	1	NA	1	2
30	Copia24-VV_LTR	2	1	NA	NA	3
31	Copia26-VV_I	NA	3	NA	115	118
32	Copia26-VV_LTR	6	14	1	5	26
33	Copia27-VV_I	NA	NA	NA	26	26
34	Copia27-VV_LTR	1	NA	NA	NA	1
35	Copia28-VV_I	NA	NA	NA	5	5
36	Copia28-VV_LTR	NA	NA	1	NA	1
37	Copia29-VV_I	NA	NA	NA	12	12
38	Copia3-VV_I	NA	NA	NA	16	16
39	Copia3-VV_LTR	4	1	6	NA	11
40	Copia4-VV_I	NA	NA	NA	13	13
41	Copia5-VV_I	NA	NA	NA	7	7
42	Copia5-VV_LTR	NA	4	NA	4	8
43	Copia6-VV_I	NA	1	NA	12	13
44	Copia6-VV_LTR	NA	2	NA	NA	2
45	Copia7-VV_I	NA	1	NA	5	6
46	Copia7-VV_LTR	1	NA	1	NA	2
47	Copia9-VV_I	NA	1	NA	12	13
48	Copia9-VV_LTR	NA	1	NA	34	35
49	Tvv1_I	NA	NA	NA	17	17
50	Tvv1_LTR	NA	2	NA	3	5
Caulimoviridae						
1	CAULIV1	NA	1	NA	20	21
LINE/L1						
1	VLINE1_VV	NA	NA	NA	46	46
2	VLINE2_VV	NA	1	NA	50	51
3	VLINE3_VV	NA	NA	NA	93	93
4	VLINE4_VV	NA	NA	NA	36	36

Table 5 Continue ...

copies in plant genes and promoter regions along with their status either they are successfully intact or truncated that was not previously identified. This study revealed several interesting features of repeats which shed new light on the plant genome organization. However, a more exhaustive analysis of repeats activity is necessary to assess their impact on host genomes due to unpredictable changes of global climatic pattern.

References

- Aaron, R., Quinlan and M.H. Ira (2010). BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics.*, **26(6)**: 841-842; <https://doi.org/10.1093/bioinformatics/btq033>.
- Bao, W., K.K. Kojima and O. Kohany (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**: 11. <https://doi.org/10.1186/s13100-015-0041-9>.
- Cheng, J., Z. Zhao, B. Li and C. Qin *et al.*, (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. *Scientific Reports.*, **6**: 18919; DOI: 10.1038/srep18919.
- Cruz, E.A.O., G.M.Q. Cruz, A.P. Vieira and M.A.V. Sluys (2016). Virus-like attachment sites as structural landmarks of plants retrotransposons. *Mobile DNA.*, **7**: 14. DOI 10.1186/s13100-016-0069-5.
- Dodsworth, S., A.R. Leitch and I.J. Leitch (2015). Genome size diversity in angiosperms and its influence on gene space. *Current Opinion in Genetics & Development.*, **35**: 73-78.
- Jaillon, O., J. Aury and B. Noel *et al.*, (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature.*, **449**: 463-467; <https://doi.org/10.1038/nature06148>.
- Krzywinski, M.I., J.E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones and M.A. Marra (2009). Circo: An information aesthetic for

Continue Table 5 ...

5	VLINE5_VV	NA	NA	NA	45	45
6	VLINE6_VV	NA	NA	NA	28	28
Harbinger						
1	Harbinger-1_VV	NA	8	NA	104	112
2	Harbinger-1N1_VV	NA	9	NA	67	76
3	Harbinger-3_VV	NA	3	NA	83	86
4	Harbinger-3N1_VV	79	59	45	84	267
5	Harbinger-3N2_VV	34	77	23	145	279
6	Harbinger-3N3_VV	NA	2	NA	7	9
7	Harbinger-3N3B_VV	NA	3	NA	1	4
8	VHARB-N2_VV	NA	53	NA	49	102
9	VHARB-N3_VV	NA	2	NA	254	256
10	VHARB4_VV	NA	19	NA	69	88
hAT						
1	hAT-10_VV	NA	NA	NA	37	37
2	hAT-10N1_VV	NA	14	NA	20	34
3	hAT-11N_VV	NA	NA	NA	23	23
4	hAT-6_VV	NA	1	NA	25	26
5	hAT-7_VV	NA	NA	NA	16	16
6	VIHAT1	NA	NA	NA	143	143
7	VIHAT2-N1_VV	NA	11	NA	51	62
8	VIHAT3-N1_VV	NA	1	NA	11	12
MuDR						
1	MuDR-11N_VV	NA	NA	NA	4	4
2	MuDR-12_VV	NA	8	NA	89	97
3	MuDR-13_VV	NA	6	NA	348	354
4	MuDR-18_VV	NA	2	NA	170	172
5	MuDR-21_VV	NA	3	NA	370	373
6	MuDR-3_VV	NA	5	NA	58	63
7	MuDR-4_VV	NA	NA	NA	47	47
8	MuDR-5_VV	NA	NA	NA	5	5
9	MuDR-6_VV	NA	5	NA	207	212
10	MuDR-7_VV	NA	NA	NA	1	1
11	MuDR-8_VV	NA	4	NA	168	172
12	MuDR-9_VV	NA	14	NA	97	111
13	MUDRAV11	NA	16	NA	337	353
14	MUDRAV12	NA	3	NA	164	167
EnSpm						
1	EnSpm-13_VV	NA	NA	NA	28	28
2	EnSpm-3_VV	NA	1	NA	12	13
3	EnSpm-4_VV	NA	8	NA	43	51
4	EnSpm-4N1_VV	NA	8	NA	44	52
5	EnSpm-5_VV	NA	14	NA	183	197
6	EnSpm-6_VV	NA	1	NA	26	27
7	EnSpm-8N_VV	NA	NA	NA	1	1
8	ENSPM-N3_VV	NA	1	NA	3	4
9	ENSPM1_VV	NA	NA	NA	10	10
10	ENSPM2_VV	NA	4	NA	18	22
Helitron						
1	Helitron1_VV	NA	NA	NA	1	1
Unclassified Transposable Element						
1	TE-7-1_VV	NA	5	NA	180	185
Unclassified DNA Transposon						
1	VHARB-N1_VV	NA	98	NA	203	301

comparative genomics. *Genome Research.*, **19(9)**: 1639-1645; <https://doi.org/10.1101/gr.092759.109>.

Macas, J., P. Novák, J. Pellicer and J. ěízková *et al.*, (2015). In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae. *PLoS ONE.*, **10(11)**: e0143424. doi:10.1371/journal.pone.0143424.

Mehrotra, S. and V. Goyal (2014). Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics Proteomics Bioinformatics.*, **12**: 164-171.

Michael, T.P. and S. Jackson (2013). The First 50 Plant Genomes. *Plant Genome* 6. doi:10.3835/plantgenome2013.03.0001in.

Naito, K., Y. Monden, K. Yasuda, H. Saito and Y. Okumoto (2014). mPing: The bursting transposon. *Breeding Science.*, **64(2)**: 109-114; doi:10.1270/jsbbs.64.109.

Negi, P., A.N.Rai and P. Suprasanna (2016). Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. *Frontiers in plant science.*, **7**: 1448; <https://doi.org/10.3389/fpls.2016.01448>.

Piégu, B., S. Bire, P. Arensburger and Y. Bigot (2015). A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution.*, **86**: 90-109; doi.org/10.1016/j.ympev.2015.03.009.

Tao, T. (2016). Standalone BLAST Setup for Unix. 2010 May 31 [Updated 2016 Sep 7]. In: BLAST® Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK52640>.

Wicker, T., F. Sabot and A. Hua-Van *et al.*, (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, **8**: 973-982; <https://doi.org/10.1038/nrg2165>.