Meta Gene 2 (2014) 694–705

Contents lists available at ScienceDirect

Meta Gene

In- silico exploration of thirty alphavirus genomes for analysis of the simple sequence repeats

Chaudhary Mashhood Alam [a], Avadhesh Kumar Singh [b], Choudhary Sharfuddin [a], Safdar Ali [b,*]

[a] Department of Botany, Patna University, Bihar 800005, India
[b] Department of Biomedical Sciences, SRCASW, University of Delhi, Vasundhara Enclave, New Delhi 110096, India

ARTICLE INFO

Article history:
Received 17 June 2014
Revised 8 September 2014
Accepted 10 September 2014
Available online 6 October 2014

Keywords:
Imperfect Microsatellite Extraction (IMEx)
Simple sequence repeats (SSR)
Compound microsatellite
Relative density
Relative abundance

ABSTRACT

The compilation of simple sequence repeats (SSRs) in viruses and its analysis with reference to incidence, distribution and variation would be instrumental in understanding the functional and evolutionary aspects of repeat sequences. Present study encompasses the analysis of SSRs across 30 species of alphaviruses. The full length genome sequences, assessed from NCBI were used for extraction and analysis of repeat sequences using IMEx software. The repeats of different motif sizes (mono- to penta-nucleotide) observed therein exhibited variable incidence across the species. Expectedly, mononucleotide A/T was the most prevalent followed by dinucleotide AG/GA and trinucleotide AAG/GAA in these genomes. The conversion of SSRs to imperfect microsatellite or compound microsatellite (cSSR) is low. cSSR, primarily constituted by variant motifs accounted for up to 12.5% of the SSRs. Interestingly, seven species lacked cSSR in their genomes. However, the SSR and cSSR are predominantly localized to the coding region ORFs for non structural protein and structural proteins. The relative frequencies of different classes of simple and compound microsatellites within and across genomes have been highlighted.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Abbreviations: SSR, Simple sequence repeat; cSSR, Compound simple sequence repeat; IMEx, Imperfect Microsatellite Extraction; RD, Relative density; RA, Relative abundance.

* Corresponding author. Tel.: +91 11 22623503; fax: +91 11 22623504.
E-mail addresses: safdar_mgl@live.in, alisafd@gmail.com (S. Ali).

http://dx.doi.org/10.1016/j.mgene.2014.09.005
2214-5400/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Introduction

According to International Committee on Taxonomy of Viruses (ICTV) report (King et al., 2012), the family *Togaviridae* has two genera namely *Alphavirus* and *Rubivirus*. *Rubivirus* genus has a solo member, *Rubella virus* whereas the *Alphavirus* includes thirty small, enveloped, plus-strand RNA viruses. These are responsible for human or animal diseases and are classified antigenically into six complexes (Garmashova et al., 2007; Powers et al., 2001). Further, the species are categorized into either the New World or Old World group based upon their prevalence.

Alphaviruses form icosahedral spherical particles of 65–70 nm in diameter and are considered a model system for structural studies of enveloped viruses. The species herein are known to cause several human diseases such as chickengunya, encephalitis and salmon pancreas diseases. Each alphavirus genome is a single-strand positive sense ~11.5Kb RNA molecule with terminal un-translated regions and, in between them, a single open reading frame (ORF) that is translated into a poly-protein. The poly-protein is cleaved after translation into four non-structural (virus replication and pathogenesis) and five structural proteins (virion synthesis) by proteinases that are a part of the polyprotein (Strauss and Strauss, 1994).

Repeat sequences have been established as highly divergent but ubiquitously present genome components across both prokaryotic and eukaryotic genomes. However, the viral genomes are now being actively pursued for their presence and significance. Tandem repeats of short motifs of DNA are called simple sequence repeats (SSR). On the basis of presence of interruptions in SSR; they are classified as interrupted, pure, compound, interrupted compound, complex and interrupted complex (Chambers and MacAvoy, 2000). These SSRs are termed compound SSR (cSSR) if two SSR are separated by a defined distance. dMAX is the maximum difference between the two SSRs of a cSSR as discussed later.

Table 1
An overview of the alphavirus genomes used for the study.

S. No	Species Id*	Name	Acc number	Genome size	GC%
1	A1	*Aura virus*	AF126284	11824	48.5
2	A2	*Barmah Forest virus*	U73745	11488	51.52
3	A3	*Bebaru virus*	HM147985	11877	48.5
4	A4	*Cabassou virus*	AF075259	11385	49.28
5	A5	*Chikungunya virus*	FN295484.2	11612	49.9
6	A6	*Eastern equine encephalitis virus*	AY722102	11680	48.71
7	A7	*Everglades virus*	AF075251	11395	49.8
8	A8	*Fort Morgan virus*	HM147986	11423	48.48
9	A9	*Getah virus*	EF631999	11689	52.3
10	A10	*Highlands J virus*	HM147988	11557	48.9
11	A11	*Madariaga virus*	EF151503	11661	48.5
12	A12	*Mayaro virus*	AF237947	11411	50.37
13	A13	*Middelburg virus*	EF536323	11674	52.6
14	A14	*Mosso das Pedras virus (78 V3531)*	AF075257	11465	49.64
15	A15	*Mucambo virus*	AF075253	11391	48.21
16	A16	*Ndumu virus*	HM147989	11688	50
17	A17	*O'nyong-nyong virus*	M20303	11835	48.32
18	A18	*Pixuna virus*	AF075256	11344	53.3
19	A19	*Rio Negro virus*	AF075258	11494	48.67
20	A20	*Ross River virus*	GQ433354	11948	51.03
21	A21	*Salmon pancreas disease virus*	AJ316244	11919	56.46
22	A22	*Semliki Forest virus*	X04129	11442	53.22
23	A23	*Sindbis virus*	HM147984	11739	51.55
24	A24	*Southern elephant seal virus*	HM147990	11245	48.63
25	A25	*Tonate virus*	AF075254	11530	49.07
26	A26	*Trocara virus*	HM147991	12052	47.81
27	A27	*Una virus*	HM147992	11964	50.5
28	A28	*Venezuelan equine encephalitis virus*	L01442	11447	49.8
29	A29	*Western equine encephalitis virus*	GQ287646	11554	48.28
30	A30	*Whataroa virus*	HM147993	11616	49.28

* The species Id given here would be used for representation throughout the manuscript.

The repeat number, length, and motif size influence microsatellite mutability. For instance, more the number of repeats; higher is the mutability (Pearson et al., 2005). Moreover, variations in copy number due to strand slippage and unequal recombination highlight the instability of the microsatellites (Tóth et al., 2000); which in turn makes it a predominant source of genetic diversity and a crucial player in viral genome evolution (Deback et al., 2009; Kashi and King, 2006). Their role in gene regulation, transcription and protein function has also been elucidated (Kashi and King, 2006; Usdin et al., 2008).

The presence and possible functional significance of SSRs in viruses have been recognized only recently (Alam et al., 2013a, 2013b, 2014a, 2014b; Chen et al., 2009, 2011, 2012; Xiangyan et al., 2011). Concerted efforts are required to identify and confirm the presence, distribution and variations of SSRs in human infecting viruses. Here, we systematically analyzed the occurrence, size, and density of different microsatellites in different species of alphaviruses, which can serve as a model for understanding functional aspects and evolutionary relationships.

Materials and methods

Genome sequences

Complete genome sequence of 30 alphaviruses species were assessed form NCBI (http://www.ncbi.nlm.nih.gov/) and analyzed for simple and compound microsatellites. Genome size of these species ranged from 11245 nt (Acc No- HM147990) to 12052 nt (Acc No- HM147991). The accession numbers, genome size and GC content of studied alphaviruses genomes have been summarized in Table 1.

Table 2
Summary of the SSRs and cSSRs observed in the studied alphavirus genomes.

S. No	Species Id	SSR	RA	RD	cSSR	cRA	cRD	cSSR%
1	A1	40	3.38	18.78	0	0.00	0.00	0.00
2	A2	58	5.05	27.68	4	0.35	5.83	6.90
3	A3	41	3.45	24.00	2	0.17	1.85	4.88
4	A4	31	2.72	19.76	2	0.18	3.16	6.45
5	A5	39	3.36	22.48	0	0.00	0.00	0.00
6	A6	33	2.83	18.92	1	0.09	0.86	3.03
7	A7	37	3.25	23.17	3	0.26	4.83	8.11
8	A8	27	2.36	17.42	1	0.09	1.58	3.70
9	A9	41	3.51	23.87	1	0.09	1.63	2.44
10	A10	41	3.55	26.56	1	0.09	1.04	2.44
11	A11	36	3.09	22.90	0	0.00	0.00	0.00
12	A12	35	3.07	21.21	2	0.18	2.98	5.71
13	A13	41	3.51	21.84	1	0.09	1.37	2.44
14	A14	32	2.79	18.58	0	0.00	0.00	0.00
15	A15	42	3.69	24.23	0	0.00	0.00	0.00
16	A16	43	3.68	26.10	1	0.09	1.71	2.33
17	A17	39	3.30	21.55	2	0.17	2.70	5.13
18	A18	52	4.58	29.62	2	0.18	2.38	3.85
19	A19	37	3.22	22.19	0	0.00	0.00	0.00
20	A20	45	3.77	29.29	2	0.17	2.93	4.44
21	A21	46	3.86	26.93	0	0.00	0.00	0.00
22	A22	32	2.80	19.31	1	0.09	1.40	3.13
23	A23	30	2.56	19.85	1	0.09	1.53	3.33
24	A24	43	3.82	27.66	1	0.09	1.16	2.33
25	A25	42	3.64	23.50	2	0.17	2.08	4.76
26	A26	37	3.07	21.99	3	0.25	4.15	8.11
27	A27	38	3.18	27.75	1	0.08	1.09	2.63
28	A28	39	3.41	23.24	1	0.09	2.53	2.56
29	A29	35	3.03	20.69	1	0.09	1.30	2.86
30	A30	28	2.41	19.63	3	0.26	4.56	10.71

Microsatellites identification

The microsatellite search was performed using the IMEx software (Mudunuri and Nagarajaram, 2007). Earlier studies on eukaryotes and *E. coli* genomes have focused on microsatellites with lengths of 12 bp or more (Tóth et al., 2000) but due to smaller size of alphaviruses genome, simple and compound microsatellite search using these parameters did not yield any results. Therefore, the search for simple and compound microsatellites was accomplished using the 'Advance- Mode' of IMEx using the parameters as used for HIV (Chen et al., 2012). The parameters were set as follows: Type of Repeat: perfect; Repeat Size: all; Minimum Repeat Number: 6, 3, 3, 3, 3, 3; Maximum distance allowed between any two SSRs (dMAX) is 10. Other parameters were used as default. Compound microsatellites were not standardised in order to determine real composition.

Statistical analysis

The simple mathematical calculations were performed using Microsoft Office Excel 2007. However, the Pearson correlation coefficient (r) was calculated using GraphPad Prism Software, version 5 (La Jolla, CA, USA) to evaluate the influence of genome size and GC content, if any, on SSRs and cSSRs. A P-value <0.05 was considered to be significant.

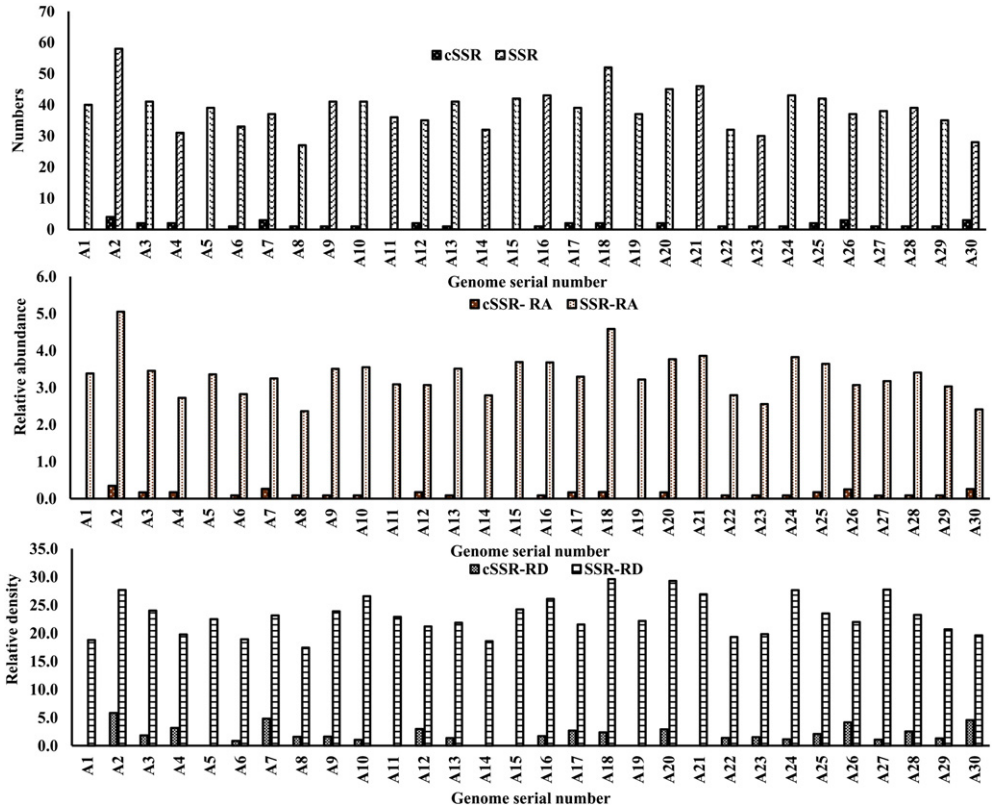

Fig. 1. Analysis of SSRs and cSSRs (a) Incidence (b) Relative abundance: SSRs/cSSRs present per Kb of genome (c) Relative density: Total length covered by SSR/cSSR per Kb of genome.

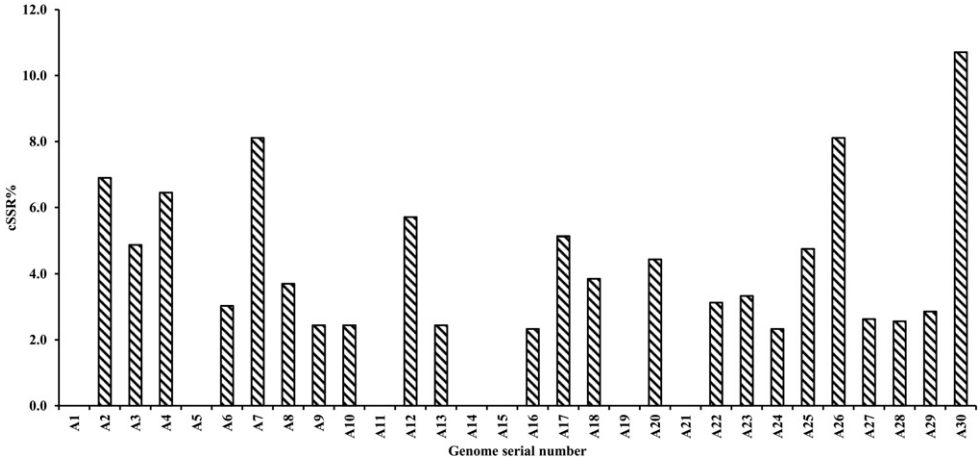

Fig. 2. Analysis of cSSR-% (Number of cSSR/Total number of SSR*100) across different alphavirus genomes.

Results

Number, relative abundance and relative density

Genome-wide scan of thirty alphavirus genomes using IMEx revealed a total of 1160 SSRs distributed across all the species with varying incident frequencies ranging from 27 in A8 (Acc No- HM147986) to 58 in A2 (Acc No- U73745) (Table 2, Supplementary Table S1 and Fig. 1a). Further, the relative abundance values lied between a minimum of 2.41 for A30 to a maximum of 5.05 bp/kb for A2. (Table 2, Fig. 1b). Comparatively, relative density of SSRs varied from 18.58 bp/kb in A14 to 29.62 bp/kb for A18 (Table 2, Fig. 1c).

IMEx scan for alphavirus genomes revealed a total of 39 cSSRs (Table 2 and Supplementary Table S2). In contrast to the ubiquitous presence of SSRs, seven alphavirus genomes lacked any cSSR. Further, relative abundance varies from 0.00 bp/kb in 7 alphavirus sequence (A1, A5, A11, A14, A15, A19, A21) to 0.35 bp/kb in A2(Acc No- U73745) (Table 2, Fig. 1b) whereas relative density varies from 0.00 bp/kb in 7 alphavirus sequence to 5.83 bp/kb in A2 (Acc No- U73745) (Table 2, Fig. 1c). The correlation of cSSR incidence in a genome to SSR abundance led to few interesting observations. cSSR% is the percentage of individual microsatellites

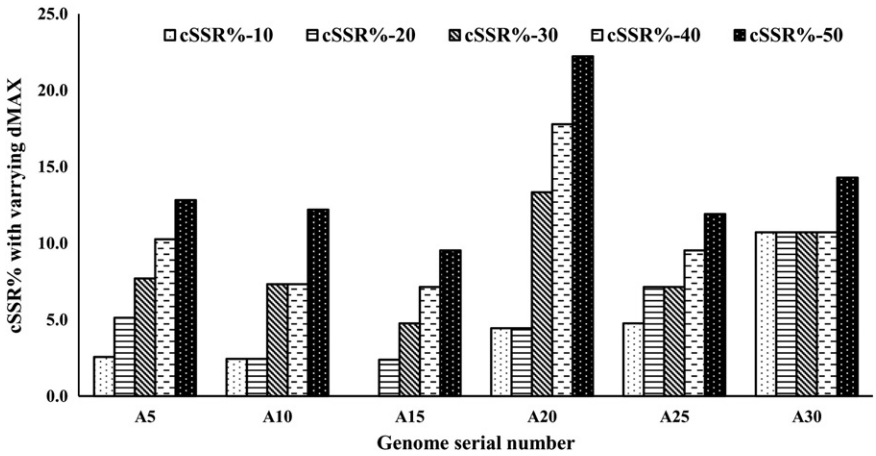

Fig. 3. Frequency of cSSR-% (Percentage of individual microsatellites being part of a compound microsatellite) in relation to varying dMAX (10 to 50) across six randomly selected alphavirus species.

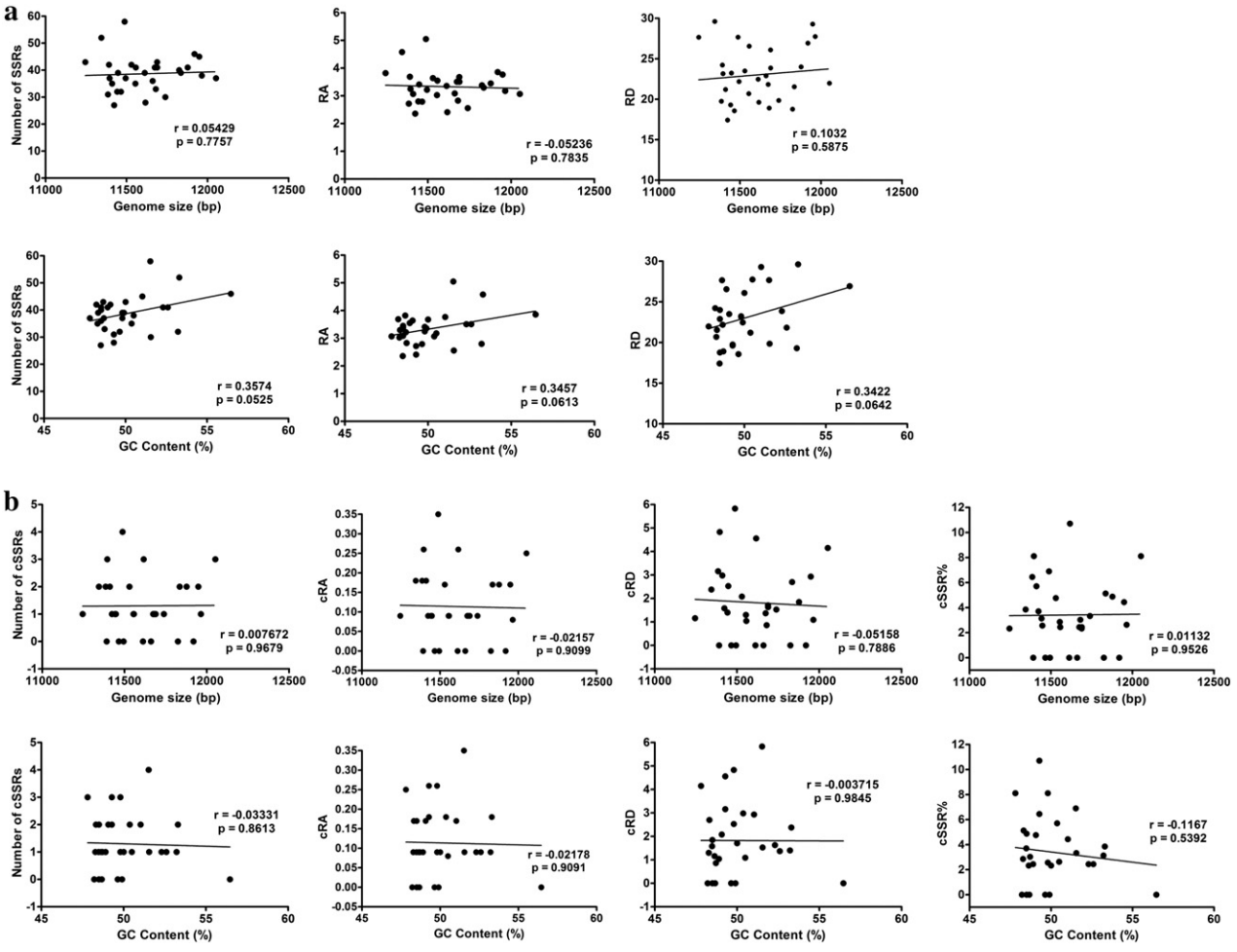

Fig. 4. Correlation analysis of (a) SSRs and (b) cSSR with genome size and GC content across alphavirus genomes.

being part of a compound microsatellite and it has been summarized in Fig. 2. It was 'nil' in C8 (Acc No- AJ620300) which was harbouring 46 SSRs as compared to 10.71% in A30 (Acc No- HM147993) with just 28 SSRs.

Varying dMAX and incidence of cSSR

dMAX is the maximum distance between any two adjacent microsatellites and if the distance separating two microsatellites is less than or equivalent to dMAX, these microsatellites are classified as cSSR (Kofler et al., 2008). To determine the impact of dMAX, compound microsatellites from A5, A10, A15, A20, A25 and A30 were chosen at random to determine the variability of compound microsatellite with increasing dMAX by analyzing cSSRs%. It is noteworthy that the dMAX value can only be set between 0 and 50 for IMEx (Mudunuri and Nagarajaram, 2007). Our analysis revealed an overall increase in cSSRs% with increasing dMAX in all the six alphavirus genomes analysed (Fig. 3).

Genome parameters and SSR/cSSR distribution

We assessed the possible influence of genome size and GC content on number/RA/RD of SSRs and cSSRs (Fig. 4; Supplementary Table S3). Genome size of assessed alphavirus genomes had a non significant influence on number of SSRs ($r = 0.05429$; $P > 0.05$), RA ($r = -0.05236$; $P > 0.05$) and RD ($r = 0.1032$; $P > 0.05$). Genome size showed similar influence on number of cSSRs ($r = 0.007672$; $P > 0.05$), and their RA ($r = -0.02157$; $P > 0.05$) and RD ($r = -0.05158$; $P > 0.05$) and cSSR% ($r = 0.001132$; $P > 0.05$). Similarly, GC content in assessed alphavirus genomes also showed non-significant correlations with number of SSRs ($r = 0.3574$; $P > 0.05$), RA ($r = 0.3457$; $P > 0.05$) and RD ($r = 0.3422$; $P > 0.05$), and also with number of cSSRs ($r = -0.03331$; $P > 0.05$), and their RA ($r = -0.02178$; $P > 0.05$) and RD ($r = -0.00371$; $P > 0.05$) and cSSR%($r = -0.1167$; $P > 0.05$).

Motifs types in analysed genomes

Mononucleotide repeats were observed in all the analyzed alphavirus genomes. Poly (A/T) repeats were significantly more prevalent (82.62%) than poly (G/C) repeats in each of the studied genomes (Supplementary Table S1) which may be attributed to the high (A/T) content of the genomes (Karaoglu et al., 2005). However, the (A/T) content being only slightly higher than G/C content in each of the analyzed sequences suggests a weak influence on the occurrence of poly (G/C) repeats. Similar to eukaryotic and prokaryotic genomes with reportedly more abundant poly (A/T) tracts (Gur-Arie et al., 2000; Karaoglu et al., 2005; Tóth et al., 2000) in the analyzed sequences poly A or poly T mononucleotide repeats prevailed over poly G or poly C (Fig. 5a).

A maximum of 64 mononucleotide A repeats was observed in genome of *Una virus* (A27). Further, the genomes had di-nucleotide repeats of six types: AG/GA, GT/TG, AC/CA, CT/TC, AT/TA, and CG/GC with variations in occurrence across genomes. AG/GA repeats were the most prevalent whereas, CT/TC was least represented (Fig. 5b). The average occurrence of GT/TG repeats was ~4 times more than CG/GC. The distribution of di-nucleotide repeats in the studied genomes has been summarized in Supplementary Table S1.

Tri-nucleotide repeats are the third most abundant SSRs within the alphavirus genomes. Of the 64 triplet repeat types, the density of AAG/GAA coding for lysine/glutamic acid was the most abundant followed by AGA coding for arginine respectively. Though differences exist in the abundance of trinucleotide repeats amongst alphavirus species but AAG/GAA showed highest prevalence in most species (Fig. 5c; Supplementary Table S1). Ten different types of tetra-nucleotide repeat motifs were distributed in 14 alphaviruses species while pentanucleotide repeat motif ATTTT was present in two genomes A7 and A27 only. There were no hexanucleotide repeats in the analysed genomes.

Comparative distribution across coding and non-coding regions

The distribution of SSRs and cSSR in coding/non-coding region revealed a clear bias of incidence in the coding regions. The coding regions accounted for 90% of the SSRs (Non-structural proteins ORF: 58%; Structural Proteins ORF: 32%) whereas residual 10% contribution was from the non-coding region of the genomes (Fig. 6a). The cSSR distribution also exhibited similar results with only 15% incidence being attributed to

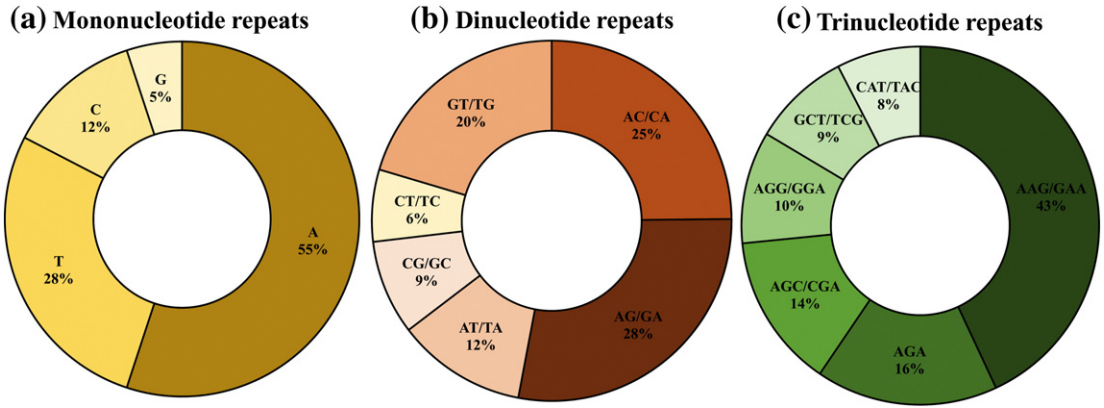

Fig. 5. Differential composition of (a) Mono-nucleotide repeats (b) Di-nucleotide repeat motifs (c) Tri-nucleotide repeat motifs.

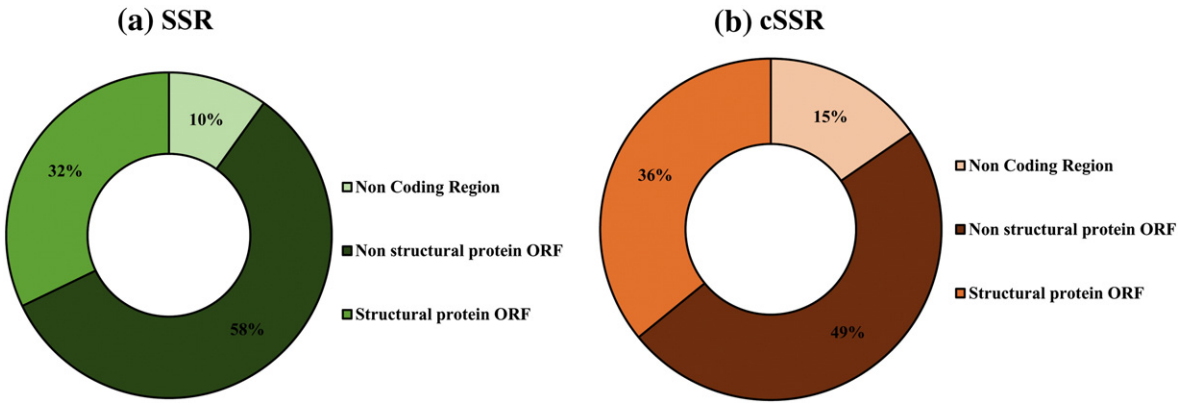

Fig. 6. Comparative distribution of (a) SSRs and (b) cSSR across coding and non-coding regions of alphavirus genomes.

Table 3
Distribution of SSRs and cSSRs across coding and non-coding regions for representative Alphavirus genomes.

S. No	Regions in genome			Number of SSRs (region)			Number of cSSRs (region)		
	Genome size (bp)	Coding (bp)	Non coding (bp)	Total	Coding	Non-coding	Total	Coding	Non-coding
A5	11612	26-7429, 7495-11241	1-25, 7430-7494, 11242-11612	39	37	2	0	0	0
A10	11557	1-5553, 7440-11150	5554-7439, 11151-11557	41	31	10	1	1	0
A15	11391	43-7410, 7446-11210	1-42, 7411-7445, 11211-11391	42	41	1	0	0	0
A20	11948	80-7561, 7606-11370	1-79, 7562–7605, 11371-11948	45	41	4	2	1	1
A25	11530	44-7549, 7585-11349	1-43, 7550–7584, 11350-11530	42	39	3	2	2	0
A30	11616	61-5625, 7526-11266	1-60, 5626-7525, 11267-11616	28	22	6	3	2	1

the non-coding regions. Their distribution across structural and non structural ORFs has been summarized in Fig. 6b. Also, the actual distribution of SSRs and cSSRs across coding and non-coding regions has been summarized in Table 3 for representative genomes.

Further, we assessed the relative composition of repeat motifs. The structural proteins ORF had predominantly trinucleotide repeats (~50%) whereas tetranucleotide motif constituted almost eight out of every ten repeats present in the non structural proteins ORF. Contrastingly, the non coding region had mononucleotide motifs as the most prevalent one (Fig. 7).

Discussion

In this study, we screened 30 alphavirus genomes for the presence, abundance, and composition of SSR tracts. A total of 1160 SSRs were revealed ranging from mono- to penta- nucleotide motifs. A higher prevalence of di-nucleotide repeats over tri-nucleotide repeats may be attributed to instability of the former because of higher slippage rate (Katti et al., 2001) and suggests a possible role of host in the evolution of di-nucleotide repeats within alphavirus (Alam et al., 2013a, 2013b, 2014a, 2014b). The repetitive sequences are hypothesized to provide for a molecular device for faster adaptation to environmental stresses (Kashi et al., 1997; Li et al., 2004) and hence, diverse repeats observed in alphavirus genomes might accelerate their evolution.

The compound microsatellites are reportedly involved in regulation of gene expression and at functional level of proteins in several species (Chen et al., 2011; Kashi and King, 2006). Though their significance in

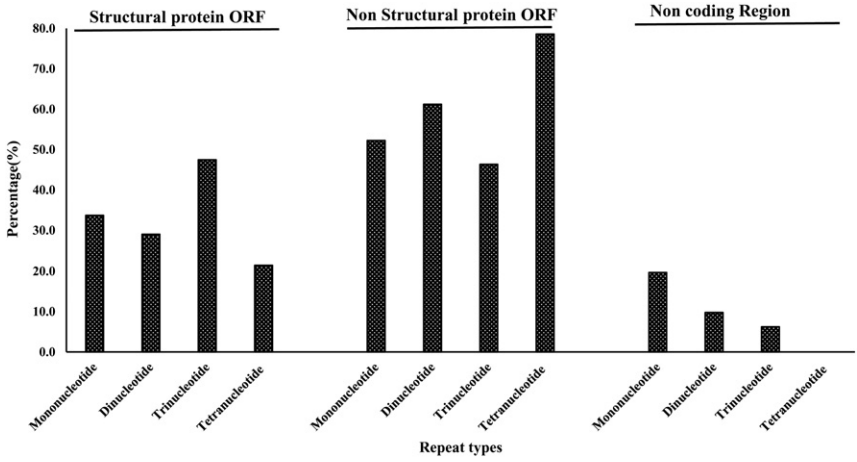

Fig. 7. Differential contribution of mono-, di- and tri-nucleotide SSR motifs across Structural proteins ORF, Non-structural proteins ORF and Non-coding regions of alphavirus genomes.

alphavirus is not clear, it suggests presence of a possibly complex regulation at the functional level. Further, varying dMAX (10 to 50) in 6 analyzed species lead to a higher cSSR% with increasing dMAX though, not in a linear way. This further adds to the significance of differential cSSRs distribution across genomes. In alphaviruses compound microsatellites were composed of a maximum of two SSRs whereas, in prokaryotes it is four and more than eight in eukaryotic species. Interestingly, cSSRs% varied between 0–11.42% in alphavirus genome; 0 to 15.5% in Human Papilloma Viruses; 0–15.15% in potyvirus genome; 0–11.43% in tobamovirus genome; 0–12.5% in carlavirus genome; 0–11.76% in potexvirus genome; (Alam et al., 2013a, 2013b, 2014a, 2014b; Singh et al., 2014), 0–24.24% in HIV-1 genomes (Chen et al., 2012); 4–25% in eight eukaryotic genomes (Kofler et al., 2008) and 1.75–2.85% in *E. coli* genomes (Kruglyak et al., 2000).

Distribution of microsatellite in the viral genome is organism specific rather than host specific. This is supported by the fact that the taxonomy of alphavirus shows no comparable congruence with host taxonomy, and species from the same lineage may have quite unrelated hosts (Gibbs et al., 2008). Accordingly, we observed that cSSR from viruses infecting common host does not possess similar number and type of cSSR motifs in their genome (data not shown). Surprisingly, seven of the alphavirus species did not possess even a single compound microsatellite this may be due to less number of strains present in these species wherein non availability of cSSR which might be restricting their variation and evolution.

It has been widely accepted that genetic RNA recombination is crucial for the emergence of new viral strains or species (Tan et al., 2005) and genomes are known to have recombination hot spots for the same (Jeffreys et al., 1998). The repeat sequences are possibly facilitating these recombination events (Nagy and Burjarski, 1996). Moreover, the unique composition of repeat motifs across structural/non structural proteins ORF and non coding regions as illustrated in Fig. 7 suggests that different parts of the genome might be evolving independent of one another in a dynalic manner.

A complete understanding of the functional and evolutionary role of tandem repeat sequences in viruses is still elusive. However, their ubiquitous presence though with varying frequency and complexity across species as well as coding and non-coding regions is suggestive of them being involved in the already established roles of gene regulation recombination hot spots. But owing to the smaller genome size of viruses as compared to prokaryotes and eukaryotes these tandem repeats are probably more influential in guiding the evolution of viruses. The diversity of microsatellites in alphavirus genomes may be useful for better understanding of viral genetic diversity, evolutionary biology, and strain/species demarcation

Conclusion

Genome-wide scan of thirty alphavirus genomes using IMEx revealed a total of 1160 SSRs and 39 cSSRs. The SSRs showed ubiquitous presence across the species albeit with varying incidence rates while seven alphavirus genomes lacked any cSSR. There was a bias towards occurrence of SSRs and cSSRs in the coding region as compared to the non-coding region which accounted for 10% and 15% contribution respectively. We postulate that such repeats in viral genomes could be involved in recombination, leading to sequence diversity that drives host adaptation. More detailed study of compound microsatellites in alphavirus genomes may be useful for understanding complex biological features such as changes in virulence and their emergence as new epidemics.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mgene.2014.09.005.

Acknowledgements

We thank Department of Biomedical Sciences, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, New Delhi, India and Department of Botany, Patna University, Bihar, India for all the financial and infrastructural support provided for the study.

References

Alam, C.M., Singh, A.K., Sharfuddin, C., Ali, S., 2014a. Incidence, complexity and diversity of simple sequence repeats across potexvirus genomes. Gene 537, 189–196 (ISSN: 0378–1119).

Alam, C.M., Singh, A.K., Sharfuddin, C., Ali, S., 2014b. Genome-wide scan for extraction and analysis of simple and imperfect microsatellites in diverse carlaviruses. Infect. Genet. Evol. 21, 287–294 (ISSN: 1567–1348).

Alam, C.M., Singh, A.K., Sharfuddin, C., Ali, S., 2013a. In silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. Gene 530, 193–200 (ISSN: 0378–1119).
Alam, C.M., George, B., Sharfuddin, C., Jain, S.K., Chakraborty, S., 2013b. Occurrence andanalysis of imperfect microsatellites in diverse potyvirus genomes. Gene 521 (2), 238–244 (1).
Chambers, G.K., MacAvoy, E.S., 2000. Microsatellites: consensus and controversy. Comp. Biochem. Physiol. B 126, 455–476.
Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G., Yu, R., 2009. Similar distribution of simple sequence repeats in diverse completed human immunodeficiency virus type 1 genomes. FEBS Lett. 583, 2959–2963.
Chen, M., Zeng, G., Tan, Z., Jiang, M., Zhang, J., Zhang, C., Lu, L., Lin, Y., Peng, J., 2011. Compound microsatellites in complete Escherichia coli genomes. FEBS Lett. 585, 1072–1076.
Chen, M., Tan, Z., Zeng, G., Zhuotong, Z., 2012. Differential distribution of compound microsatellites in various human immunodeficiency virus type 1 complete genomes. Infect. Genet. Evol. 12, 1452–1457.
Deback, C., Boutolleau, D., Depienne, C., Luyt, C.E., Bonnafous, P., Gautheret-Dejean, A., Garrigue, I., Agut, H., 2009. Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. J. Clin. Microbiol. 47, 533–540.
Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E., Frolov, I., 2007. The old world and new world alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. J. Virol. 81 (5), 2472–2484.
Gur-Arie, R., Cohen, C.J., Eitan, Y., 2000. Simple sequence repeats in Escherichia coli: abundance, distribution, composition, and polymorphism. Genome Res. 10, 62–71.
Gibbs, A.J., Ohshima, K., Phillips, M.J., Gibbs, M.J., 2008. The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. PLoS One 3, e2523.
Jeffreys, J., Murray, J., Neumann, R., 1998. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol. Cell 2, 267–273.
Karaoglu, H., Lee, C.M., Meyer, W., 2005. Survey of simple sequence repeats in completed fungal genomes. Mol. Biol. Evol. 22, 639–649.
King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2012. Virus taxonomy: classification and nomenclature of viruses. Ninth Report of the International Committee on Taxonomy of Viruses, San Diego.
Kashi, Y., King, D., Soller, M., 1997. Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13, 74–78.
Kashi, Y., King, D.G., 2006. Simple sequence repeats as advantageous mutators in evolution. Trends Genet. 22, 253–259.
Katti, M.V., Ranjekar, P.K., Gupta, V.S., 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. 18, 1161–1167.
Kofler, R., Schlotterer, C., Luschutzky, E., Lelley, T., 2008. Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. BMC Genomics 9, 612.
Kruglyak, S., Durrett, R., Schug, M.D., Aquadro, C.F., 2000. Distribution and abundance of microsatellites in the Yeast genome can be explained by a balance between slippage events and point mutations. Mol. Biol. Evol. 17, 1210–1219.
Li, Y.C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. 21, 991–1007.
Mudunuri, S.B., Nagarajaram, H.A., 2007. IMEx: imperfect microsatellite extractor. Bioinformatics 23, 1181–1187.
Nagy, P.D., Burjarski, J.J., 1996. Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. J. Virol. 70, 415–426.
Pearson, C.E., Nichol Edamura, K., Cleary, J.D., 2005. Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet. 6, 729–742.
Powers, A.M., Brault, A.C., Shirako, Y., Strauss, E.G., Kang, W., Strauss, J.H., Weaver, S.C., 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. 75 (21), 10118–10131.
Singh, A.K., Alam, C.M., Sharfuddin, C., Ali, S., 2014. Frequency and distribution of simple and compound microsatellites in forty-eight Human Papillomavirus (HPV) genomes. Infect. Genet. Evol. 24, 92–98.
Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 58 (3), 491–562 (Review. Erratum in: Microbiol Rev 58(4),806).
Tan, Z., Gibbs, A.J., Tomitaka, Y., Sanchez, F., Ponz, F., Ohshima, K., 2005. Mutations in Turnip mosaic virus genomes that have adapted to Raphanus sativus. J. Gen. Virol. 86, 501–510.
Tóth, G., Gáspári, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10, 967–981.
Usdin, K., 2008. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. Genome Res. 18, 1011–1019.
Xiangyan, Z., Zhongyang, T., Haiping, F., Ronghua, Y., Mingfu, Li, Jianhui, J., Guoli, S., Ruqin, Y., 2011. Microsatellites in different Potyvirus genomes: survey and analysis. Gene 488, 52–56.