BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 17(10)

Codon usage signatures in Sabia and Chapare for host adaptation

Himani Malhotra[1]* & Arvind Kumar[2]*

Department of Biotechnology[1], Department of Biochemistry[2], School of Bioengineering and Biosciences, Lovely Professional University, Jalandhar Delhi G.T. Road, Phagwara, Punjab, INDIA -144411. *Corresponding author: Arvind Kumar – E-mail: arvind.19345@lpu.co.in; Phone:+91-9815289975; Himani Malhotra - E-mail: himani.11814490@lpu.in,Tel:+91-9463133277

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:
Sabia and Chapare viruses in the Arenavirus family cause viral hemorrhagic fever among humans with a fatality rate of 30% with no treatment models. Therefore, it is of interest to document the codon usage, amino acid patterns and associated factors influencing the observed variations in Sabia and Chapare viruses for host adaptation. Multivariate statistical analysis revealed compositional constraint and host selection pressure influencing the viral codon usage patterns. These data suggests the codon usage signatures in Sabia and Chapare viruses for host adaptation in the human host implying its role in the rapid progression of the infection. Dinucleotides UpG and CpA were noted to be over-represented among the Sabia, Chapare viruses and human genomes. Strong restraint from the usage of CpG dinucleotides among viruses is linked with the molecular mimicry of the human immune system. Thus, the data reported from this study help in understanding the mechanism of viral adaptation inside the host genome for further consideration in drug discovery.

Keywords: Sabia virus; Chapare virus; *Homo sapiens*; codon usage; dinucleotides

Background:
Study of viruses to eradicate them globally is of great concern considering their highly detrimental association with all type of life forms including bacteria, archaea, and eukaryotes (human and agricultural sector, zoonotic threats) [1]. *Arenaviridae* represent family of viruses disseminating diseases among humans through direct or indirect contact with rodents [2]. Infections caused by Arenaviruses prevail more commonly in areas of South America, South Africa and have been described to be federated with severe disturbance among humans [3]. *Arenaviridae* mainly comprises of three genera *Mammarenavirus*, *Reptarenavirus* and *Hartmanivirus*

including viruses infecting mammals, reptiles and fishes [4].The genome of Arenaviruses possessing negative sense single-stranded RNA encompasses two segments pertaining Small (S) RNA segment of size 3.4 kb encoding for envelope glycoprotein precursor (GPC) and the nucleoprotein (NP);Large (L segment) of size 7.2 kb encoding for matrix protein (Z) as well as the viral RNA-dependent RNA polymerase protein(L) [5]. On the basis of similarity in geographical distribution, antigenic properties and also on phylogenetic data genus *Mammarenavirus* have been subdivided into Old World Arenaviruse (OW) and New World Arenaviruse (NW)[7].Subgroups of OW and NW Arenaviruses

include total 10 strains causing diseases among humans and are also examined as polyphyletic [6, 7].Further New World Arenaviruse have been sub grouped into clades: A, B, C and D. Five viruses of clade B of NW Arenaviruse; known to be pathogenic among humans are Junin, Machupo, Guanarito, Sabia and Chapare [8]. Clade B viruses have been symbolized as an emergence among humans due to their categorization as type A pathogen and menace as a bioterrorism agent [9].

Sabia virus causing Brazilian haemorrhagic fever was first isolated in Sao Paulo in 1994 and Chapare virus causing Bolivian haemorrhagic fever was first isolated in Chapare Province in 2003 [10, 11]. Yellow fever was the initial suspicion in case of the Sabia virus and also correlated with the Chapare virus infection as both had identified extensive liver necrosis [11]. The rodent host species for both the viruses are still unknown [10]. Apart from pervasion of studies for identification of therapeutic facilities for prevention and cure of Sabia and Chapare virus, no drug out of date being administered [4]. The availability of genomic sequencing data sprouted ample opportunities to study the riddles of the viruses at genomic level and to explore the convoluted methods showing that these viruses infect their host [12]. Therefore, study of synonymous codons that are considered to be equivalent and interchangeable has shown that alteration in synonymous codons affect the protein biogenesis which includes transcription, translation, posttranslational modifications, co translational modifications, hydrophobicity, hydrophilicity, the secondary structure of proteins, the abundance of tRNA and interaction between codon and anticodons [13-15].

Viral genomes, depends on the host machinery and cellular microenvironment for protein biogenesis, survival and progression of infection so this influences the requirement for exploration of viral host codon usage patterns [12, 16]. Deciphering the variations and factors regulating the complicated patterns of codons and amino acids of viral genome may stimulate information regarding the regulation of host by viruses which may be utilized to design therapeutics and vaccines against virus with high accuracy [17]. Therefore, it is of interest to document the codon usage, amino acid patterns and associated factors influencing the observed variations in Sabia and Chapare viruses for host adaptation.

Materials and Methods:
Retrieval of Data:
Whole coding sequences of Sabia and Chapare viruses (Table 1) were downloaded from GenBank [18] and Virus Pathogen Resource database [19]. Coding sequence of *H.sapiens* (GRCh38.p13), common host to both the viruses was also extracted from GenBank [18] for further investigation (Figure 0).

Assessment of parameters pertaining to Nucleotide composition and codon usage analysis:
Nucleotide composition properties like %A (Adenine),%G (Guanine), %C(Cytosine) and %T(Thymine); occurrence of GC (Guanine + Cytosine) at all the three positions of synonymous codons (GC1, GC2 and GC3); overall occurrence of AT and GC in

Sabia and Chapare viruses were examined using CAIcal server [20]. RSCU (Relative synonymous codon usage) was computed by CodonW (Ver. 1.4.2) software [21].Codons having RSCU greater than 1.0 demonstrate positive codon usage biasness.
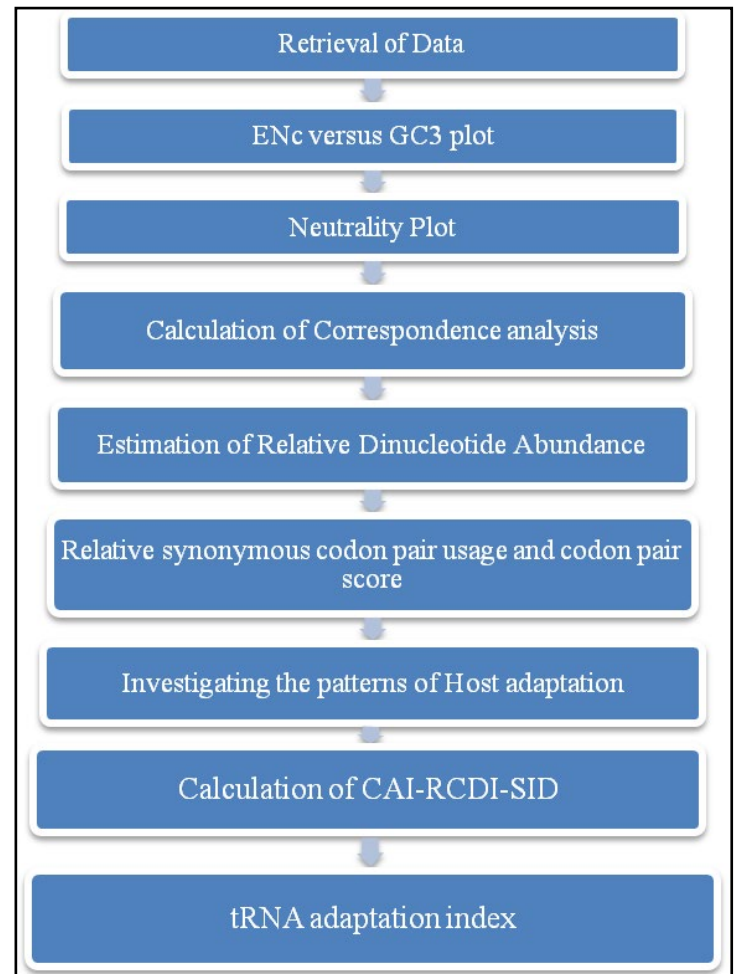

Figure 0: Graphical abstract of the work

Effective number of codons:
Effective number of codons (ENc) computed from CodonW [21] can have values from 20 to 61.Value equal to or close to 20 depicts that each amino acid has been encoded by one single codon only and there is no biasness whereas, value equal to or close to 61 shows that a particular amino acid can be encoded by more than one codon which is the case with no codon biasness. Codon usage patterns were computed by plotting ENc-GC3 plot [19].

Neutrality plot:
Neutrality plot provides information about effect of mutational constraints and natural selection on genes of viral genome. Slope value of the regression line (close to or above 1) reflects the consequence of mutational constraint only, value (close to or below 0) reflects natural selection effect also [22].

©Biomedical Informatics (2021)

Correspondence analysis (CoA) of codon and amino acid usage data:

Correspondence analysis with a p-values less than or equal to 0.05 and 0.01 was performed using SPSS (Statistical Package for the Social Sciences) software to depict the changes in patterns of codon and amino acid in genome sequence [17,23].

Estimation of Relative Dinucleotide Abundance

Relative Dinuclotide Abundance (P_{xy}) was analyzed using CAIcal server [37]. P_{xy} value greater than 1.25 depicts over-representation of dinucleotides and P_{xy} value less than 0.78 show under-representation of dinucleotides [24].

Computation of Codon Pair Score and Relative Synonymous Codon Pair Usage

Relative Synonymous Codon Pair Usage (RSCPU) represented as ratio of observed frequencies to the expected frequencies of codon pairs. RSCPU values were computed by using an in-house BioPerl script and further RSCPU values are used to analyze the Codon pair Score (CPS) values for codon pairs of Sabia and Chapare viruses and its host human by using script. Positive CPS scores show over-representation of codon pairs, whereas, negative CPS scores depicts under-representation of codon pairs for virus and host [25].

Codon adaptation index (CAI)

Values of CAI computed by CAIcal server ranges from 0 to 1 estimate the adaptation of viral genes inside the host cellular environment by using set of highly expressed reference genes. High CAI value (close to 1) of a concerned gene indicates immense level of similarity in its codon usage pattern with host and tremendous adaptation in host environment [17].

Relative codon deoptimization index

Relative codon deoptimization index (RCDI) analyzes the degree of acclimatization of viral genomes in host microcellular environment and were assessed by RCDI/eRCDI server [26].If RCDI value is low indicating better adaptation and increased translation of a viral gene segment in host system [27].

Similarity index

Similarity index estimates the magnitude of the impact of host genome in driving codon usage patterns of viruses. Similarity index values ranges from 0 to 1, value close to 1 implies a thorough effect of host on viral codon usage [28].

Examination of tRNA adaptation index

tRNA adaptation index (tAI)estimates usage of tRNA by the coding sequences of viral genome. tAI defines adaptation level of coding sequence of virus with the corresponding tRNA pool of host cell by computing the presence of tRNAs for every codon of coding sequence [29].

Table1: Relative synonymous codon usage analysis of Sabia virus and *Homo sapiens*

Codon(Aa)	Sabia	*Homo sapiens*	Codon(Aa)	Sabia	*Homo sapiens*
TTT(Phe)*	1.27	0.92	GCT(Ala)	1.15	1.08
TTC(Phe)	0.73	1.08	GCC(Ala)	0.62	1.60
TTA(Leu)	1.28	0.48	GCA(Ala)*	2.13	0.92
TTG(Leu)*	1.65	0.78	GCG(Ala)^	0.10	0.44
CTT(Leu)	1.12	0.78	TAT(Tyr)*	1.19	0.88
CTC(Leu)^	0.53	1.20	TAC(Tyr)	0.81	1.12
CTA(Leu)^	0.59	0.42	CAT(His)*	1.13	0.84
CTG(Leu)	0.82	2.40	CAC(His)	0.87	1.16
ATT(Ile)*	1.27	1.08	CAA(Gln)*	1.20	0.54
ATC(Ile)	0.71	1.41	CAG(Gln)	0.80	1.46
ATA(Ile)	1.02	0.51	AAT(Asn)*	1.31	0.94
GTT(Val)*	1.35	0.72	AAC(Asn)	0.69	1.06
GTC(Val)^	0.56	0.96	AAA(Lys)*	1.09	0.86
GTA(Val)	0.79	0.48	AAG(Lys)	0.91	1.14
GTG(Val)	1.30	1.84	GAT(Asp)*	1.29	0.92
TCT(Ser)	1.58	1.14	GAC(Asp)	0.71	1.08
TCC(Ser)	0.69	1.32	GAA(Glu)*	1.21	0.84
TCA(Ser)*	1.81	0.90	GAG(Glu)	0.79	1.16
TCG(Ser)^	0.17	0.30	TGT(Cys)*	1.54	0.92
AGT(Ser)	1.13	0.90	TGC(Cys)^	0.46	1.08
AGC(Ser)	0.63	1.44	CGT(Arg)^	0.26	0.48
CCT(Pro)	1.42	1.16	CGC(Arg)^	0.13	1.08
CCC(Pro)	0.71	1.28	CGA(Arg)^	0.25	0.66
CCA(Pro)*	1.51	1.12	CGG(Arg)^	0.12	1.20
CCG(Pro)^	0.36	0.44	AGA(Arg)*	3.23	1.26
ACT(Thr)	1.23	1.00	AGG(Arg)	2.02	1.26
ACC(Thr)	0.87	1.44	GGT(Gly)*	1.48	0.64
ACA(Thr)*	1.79	1.12	GGC(Gly)	0.60	1.36
ACG(Thr)^	0.10	0.44	GGA(Gly)	1.06	1.00
			GGG(Gly)	0.86	1.00

Aa stands for Amino acids; Codons having RSCU (Relative synonymous codon usage) > 1.00 have been marked in bold; Codons rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; Highly preferred codons for each amino acid has been marked with *; Under-represented codons having RSCU value less than 0.60 has been marked with ^; Codons showing richness in G(Guanine) or C(Cytosine) nucleotides have been highlighted in green.

Table2: Relative synonymous codon usage analysis of Chaparevirus and *Homo sapiens*

Codon (Aa)	Chapare	*Homo sapiens*	Codon(Aa)	Chapare	*Homo sapiens*
TTT(Phe)*	1.24	0.92	GCT(Ala)*	1.55	1.08
TTC(Phe)	0.76	1.08	GCC(Ala)	0.90	1.60
TTA(Leu)	1.31	0.48	GCA(Ala)	1.46	0.92
TTG(Leu)*	1.38	0.78	GCG(Ala)	0.09	0.44
CTT(Leu)	1.04	0.78	TAT(Tyr)*	1.25	0.88
CTC(Leu)	0.67	1.20	TAC(Tyr)	0.75	1.12
CTA(Leu)	0.72	0.42	CAT(His)*	1.14	0.84
CTG(Leu)	0.89	2.40	CAC(His)	0.86	1.16
ATT(Ile)*	1.27	1.08	CAA(Gln)*	1.12	0.54
ATC(Ile)	0.83	1.41	CAG(Gln)	0.88	1.46
ATA(Ile)	0.90	0.51	AAT(Asn)*	1.14	0.94
GTT(Val)*	1.41	0.72	AAC(Asn)	0.86	1.06
GTC(Val)	0.80	0.96	AAA(Lys)*	1.18	0.86
GTA(Val)	0.53	0.48	AAG(Lys)	0.82	1.14
GTG(Val)	1.26	1.84	GAT(Asp)*	1.20	0.92
TCT(Ser)	1.34	1.14	GAC(Asp)	0.80	1.08
TCC(Ser)	0.70	1.32	GAA(Glu)*	1.16	0.84
TCA(Ser)*	1.73	0.90	GAG(Glu)	0.84	1.16

©Biomedical Informatics (2021)

TCG(Ser)^	0.25	0.30	*TGT(Cys)**	1.27	0.92
AGT(Ser)	1.29	0.90	TGC(Cys)	0.73	1.08
AGC(Ser)	0.70	1.44	CGT(Arg)^	0.19	0.48
CCT(Pro)	1.20	1.16	CGC(Arg)^	0.23	1.08
CCC(Pro)	0.80	1.28	CGA(Arg)^	0.19	0.66
CCA(Pro)*	1.71	1.12	CGG(Arg)^	0.11	1.20
CCG(Pro)	0.29	0.44	*AGA(Arg)**	3.28	1.26
ACT(Thr)	1.21	1.00	AGG(Arg)	2.00	1.26
ACC(Thr)^	0.47	1.44	**GGT(Gly)***	1.51	0.64
*ACA(Thr)**	2.11	1.12	GGC(Gly)^	0.52	1.36
ACG(Thr)^	0.21	0.44	GGA(Gly)	1.14	1.00
			GGG(Gly)	0.84	1.00

Aa stands for Amino acids; Codons having RSCU (Relative synonymous codon usage) value >1.00 have been marked in bold; Codons rich in A (Adenine) or T (Thymine) nucleotides have been marked in red; Highly preferred codons for each amino acid has been marked with *; Under-represented codons having RSCU value less than 0.60 has been marked with^; Codons showing richness in G (Guanine) or C (Cytosine) nucleotides have been highlighted in green.

Table3: Showing average values of nucleotides in genome of virus

Organisms	%A	%U	%G	%C	%AU	%GC	%AU3	%GC3
Sabia	32.02	27.38	19.407	21.18	59.406	40.59	58.82621	41.17379
Chapare	32.61	25.64	20.84	20.88	58.26	41.73	58.10367	41.89633

Average values of nucleotides of viral genome show preference for AU rich codons as percentage of AU is much higher than GC and also AU3 is preferred over GC3 showing preference of AU also at third position of codon.

Table 4: Correlation analysis of various parameters with the Axis 1 and 2 of RSCU data

Organism		GC	GC3	RCDI	CAI	Length	GRAVY	Aromo
Sabia	Axis1 (RSCU)	-0.452	-0.539	-.855**	-.952**	0.572	0.414	0.607
	Axis 2 (RSCU)	.633*	0.203	0.045	-0.259	-0.178	-.779**	-.789**
		GC	GC3	RCDI	CAI	Length	GRAVY	Aromo
Chapare	Axis1 (RSCU)	-0.452	-0.539	-.855**	-.952**	0.572	0.414	0.607
	Axis 2 (RSCU)	.633*	0.203	0.045	-0.259	-0.178	-.779**	-.789**

**symbol shows statistically significant results at P-value less than 0.01; *symbol depicts statistically significant results at P-value less than 0.05; RSCU stands for Relative synonymous Codon usage; Length shows Length of protein sequences; GRAVY shows grand average hydropathicity score of proteins; Aromo depicts aromaticity of encoded proteins; CAI shows the codon adaptation index; RCDI stands for relative codon deoptimization index.

Table 5: Analysis of preferred codons in Sabia virus and iso-acceptor tRNAs in *Homo sapiens*.

Amino acids	Most preferred codons in Sabia virus	tRNA isotypes in Homo sapiens
Ala	GCA	AGC (22), GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), **UCC (9)**
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTT	**AAC (9)**, GAC (0), CAC (11), UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGA	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), **UCU (6)**
Leu	TTG	AAG (9), GAG (0), CAG (9), UAG (3),**CAA (6),** UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0), GUC (13)
Glu	GAA	CUC (8), **UUC (7)**
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14)**, GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)
Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *Homo sapiens* matching the most preferred; codons of Sabia virus are marked in bold.

Table 6: Analysis of highly preferred codons in Chapare virus and iso-acceptor tRNAs in *Homo sapiens*

Amino acids	Most preferred codons in Chapare virus	tRNA isotypes in Homo sapiens
Ala	GCT	**AGC (22),** GGC (0), CGC (4), UGC (8)
Gly	GGT	ACC (0), GCC (14), CCC (5), **UCC (9)**
Pro	CCA	AGG(9),GGG(0),CGG(4),**UGG(7)**
Thr	ACA	AGU (9), GGU (0), CGU (5), **UGU (6)**
Val	GTG	AAC (9), GAC (0), **CAC (11)**, UAC (5)
Ser	TCA	AGA (9), GGA (0), CGA (4), UGA (4),ACU (0), GCU (8)
Arg	AGA	ACG (7), GCG (0), CCG (4), UCG (6), CCU (5), **UCU (6)**
Leu	TTG	**AAG (9),** GAG (0), CAG (9), UAG (3),CAA (6), UAA (4)
Phe	TTT	AAA (0), GAA (10)
Asn	AAT	AUU (0), GUU (20)
Lys	AAA	CUU (15), **UUU (12)**
Asp	GAT	AUC (0), GUC (13)
Glu	GAA	CUC (8), **UUC (7)**
His	CAT	AUG (0), GUG (10)
Gln	CAA	CUG (13), UUG (6)
Ile	ATT	**AAU (14),** GAU (3), UAU (5)
Tyr	TAT	AUA (0), GUA (13)

Cys	TGT	ACA(0),GCA(29)

Most abundant iso-acceptor tRNAs in *Homo sapiens* matching the most preferred codons of Chapare virus are marked in bold.

Results and Discussion:

Through assessment of RSCU data it was inspected that out of all possible codon sets (excluding start and stop codons) as shown in Table1 and 2; 49.45% in Sabia and 47.45% in Chapare were preferred (RSCU greater than 1.0) codon sets respectively.Extensive analysis of genomic composition in the present study revealed that AU rich codons show preference over GC rich codons in Sabia and Chapare viruses shown in Table3. It was also perceptible from robust codon usage analysis as in Table1 and Table2 that Sabia and Chapare viruses had low codon usage biasness. Similar cases of RNA viruses showing low codon usage biasness have been reported earlier also [17, 30 and31]. Low codon usage biasness in viral genome reduces the competition of the virus with its host for usage of host machinery for synthesis and increases the efficiency of replication and easy adaptation inside the host cells [17, 32].
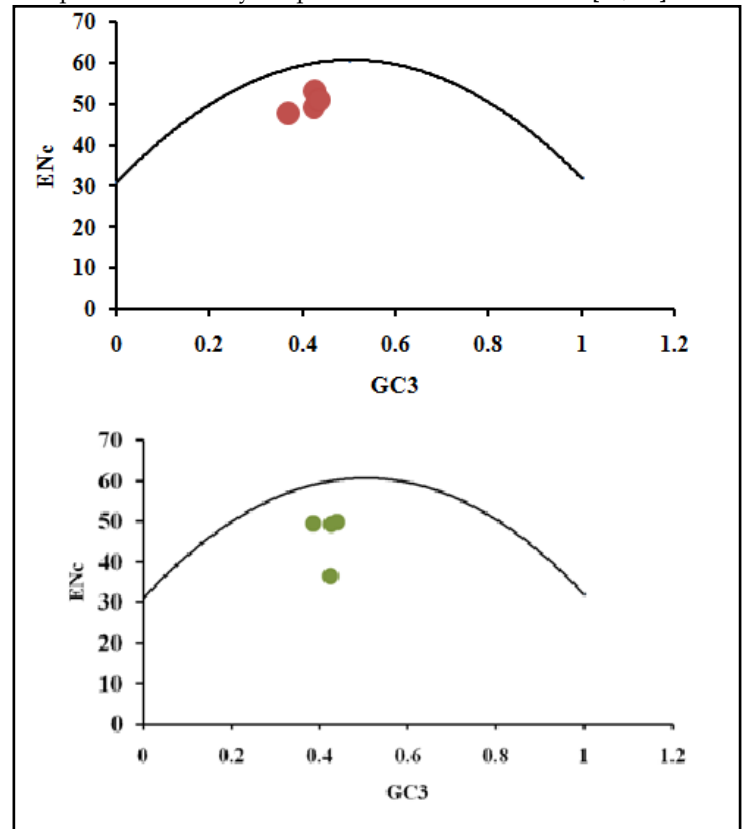

Figure 1: GC3 versus ENc plot for a) Sabia virus b) Chapare virus. Examined viral genes are being marked with red color in Sabia virus and green color in Chapare virus.

Parameters affecting codon usage data were inferred from ENc versus GC3 plots and Neutrality plot [33]. If viral gene values prevail above or fall on the curve, mutational biasness is the only aspect affecting the codon usage. However, values lying below the curve signify the occurrence of natural selection also. In-depth study of the ENc versus GC3plot (Figure 1a and 1b) of Sabia and

Chapare viral nucleotide sequences revealed the clustering of viral genes below the ENc curve. Such an observation illustrated the integrated impact of mutational constraint and evolution on codon usage patterns of Sabia and Chapare genomes. Average ENc values were found to be 50.144 ± 2.07 for Sabia and 46.2375 ± 6.038 for Chapare virus. However, analysis of neutrality plot of Sabia and Chapare viruses revealed (Figure 2a and 2b) 0.692, 0.821 slope of regression line signifying 69.2% and 82.1% impact of mutational pressure. Thus, it was evident that the effect of compositional constraint has been stronger than natural selection [33]. Further, Correspondence analysis was executed to classify the determinants causing variation in codon usage. Immense level of significant correlation of GC with Axis2 (one of the major axis of separation of genes) of RSCU data was observed in Sabia and Chapare viruses showing the influence of compositional constraint (Table 5).
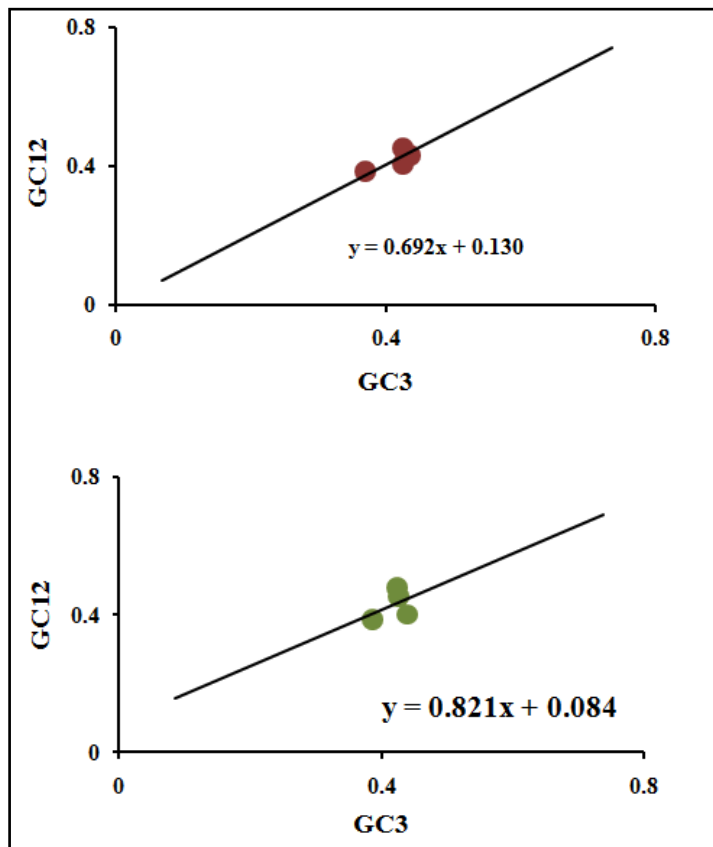
Figure 2: Neutrality plot of a) Sabia b) Chapare virus. Inspected viral genes have been marked as red coloured circles in Sabia virus and green coloured circles in Chapare virus. Slope of the plot depicts the degree of compositional bias operative on the genomes of interest

RSCU data on Axis1 commence to show significant correlation with CAI, RCDI of Sabia and Chapare viral genomes (Table 5), thus, analyzing an indubitable affect of natural selection. Elements such as GRAVY (grand average of hydropathicity) and aromaticity show significant level of correlation with RSCU data on Axis2. Thus, codon usage patterns of the Sabia and Chapare viruses found to be

a complex interplay of diverse crucial determinants. This analysis predicts that codon usage patterns of both Sabia and Chapare viruses found to be afflicted by many factors like mutational biasness, natural selection; hydropathicity and aromaticity [34, 35].Yet, in spite of a convoluted interplay of various determinants, compositional constraint was found to play the most dominant role in shaping codon usage of Sabia and Chapare viruses.

Further vigorous analysis of relative dinucleotide abundance in Sabia and Chapare viruses revealed that UpG and CpA dinucleotides were over-represented and dinucleotide CpGs, were found to be under-represented among Sabia and Chapare viral genome (Figure3 (a, b)). Similar patterns of dinucleotides were also observed to be highly preferred in *H. sapiens* also. Dinucleotides have a great influence on codon usage pattern and such feature of under-representation of CpGs dinucleotide has been observed in various genomes of RNA viruses [36]. It has been proposed that coding sequences of viral pathogens having unmethylated CpG have been recognized as pathogen signature's by host receptor Toll like receptor 9 (TLR9) and stimulates innate immune responses in host(human)[37]. However, presence of under-representation of CpGs dinucleotide will decline the host immune response and bring about increase in viral infection among host. Also, analysis of viral genome data in our study proved that selective pressure with evolution has influenced the dinucleotide pattern and also codon usage of humans.
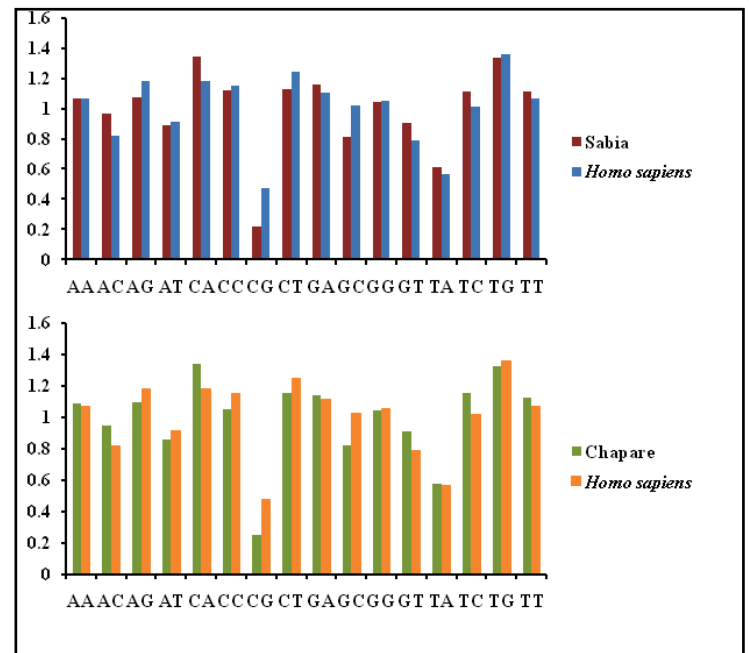

Figure 3: Relative Dinucleotide analysis of a) Sabia b) Chapare virus. X-axis showing Dinucleotides and legends on right showing name of virus, host: *Homo sapiens*
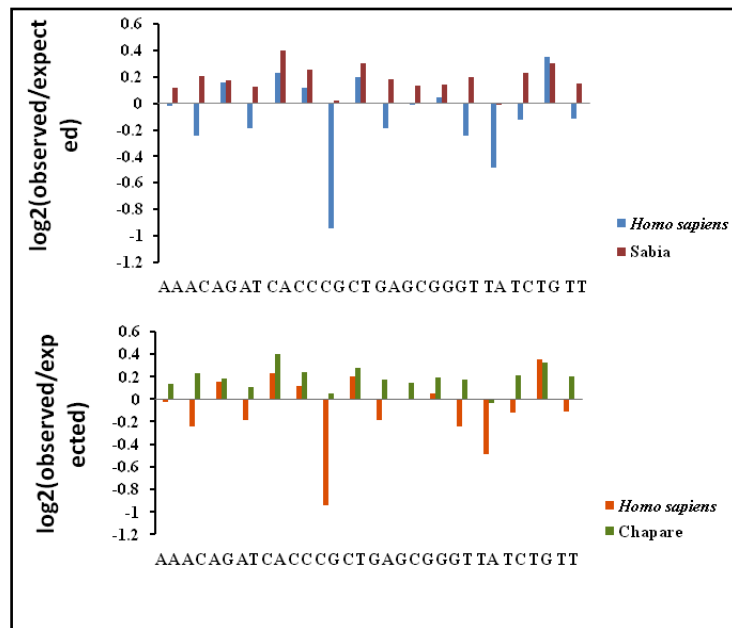
Figure 4: CPS (Codon pairing Score) results of a) Sabia b) Chapare virus. X-axis showing Dinucleotides and legends on right showing name of virus, host: human.

Extensive analysis of Sabia virus codon pairs reported that based on RSCPU values; 1237 out of total 3721 codon pairs (excluding stop:stop and stop:sense codon pairs)were found to be over-represented and 519 were under-represented. GCG-ACC codon pair coding for Alanine-Threonine was utmost over-represented and codon pair ACA-AAG coding for Threonine-Lysine as shown in Table1 was utmost over-represented. Interestingly, in Sabia virus where 55.5% matched with the over-represented codon pairs in *H. sapiens*. Similar trend was also evident among under-represented codon pairs where 48.16% matched with that of the under-represented codon pairs of the human genome.

Similarly, thorough study of RSCPU values of Chapare virus explained that 1249 out of 3721were found to be over-represented, 533 were under-represented. CGG-CCC codon pair coding for Arginine-Proline was utmost over-represented and codon pair UUC-GAG, encoding for Phenylalanine-Glutamate pair as in Table 2,was examined as utmost under-represented in Chapare virus. Interestingly, in Chapare virus 56.6% matched with the over-represented codon pairs and 47.65% matched with that of the under-represented codon pairs of the human genome.

Similar trend was also evident among under-represented viral codon pairs as 254 out of 533 (Dinucleotide pattern NNU-GNN (UpG dinucleotide) was depicted as one of the most prevalent (10.6% in Sabia virus and 11.04% in Chapare virus) as compare to the other over-represented codon pairs (Figure 4a, b). In addition, methodical inspection at the codon pair interface (cP3-cA1) determined that UpG, CpA, and CpU dinucleotides, were prevalent at the codon-codon junctions in Sabia and Chapare viruses (Figures 4 a, b). Interestingly, exactly same dinucleotide

patterns were also noted to be predominant among the codon pairs in *H. sapiens*, revealing efficient adaptation of viruses in humans.

Sabia and Chapare viruses were found to display antagonism with human host (Table1 and2). Past study revealed that antagonistic codon patterns decreases the translational efficacy but leads to proper and correct folding of viral proteins. Various parameters such as Codon adaptation index, Relative codon deoptimization index and Similarity index of viral genes analyzed the adaptation of viruses among host *Homo sapiens.* The average value of Codon adaptation index of Sabia virus was 0.76±0.03 and Chapare virus was 0.75±0.02.The average RCDI value of Sabia virus was 1.40±0.04 and Chapare virus was 1.41±0.23. The SiD values computed for the Sabia virus and Chapare virus was 0.072 and 0.073 showing the low impact of human host on viral codon biasness. These results predict high level of adaptation of viruses in *H.sapiens* [17, 25].

Examination of highly favoured codons in Sabia, Chapare viruses and isoacceptor tRNAs present in human cells divulged that 9 codons out of 18(Table5) highly favoured codons in Sabia virus;10 out of 18 in Chapare virus(Table6) correspond together with the relevant isoacceptor tRNAs present in human hosts. On the whole the highly preferred codons examined in viral coding sequences utilize suboptimal isoacceptor tRNAs present in human cells (Table 5 and 6). Similar results have also been reported for Nipah virus to recognize the usage of suboptimal tRNA isotype. It has been proposed that throughout the initial phase of an infection; the utilization of suboptimal isoacceptor host tRNAs might lead to gradual and exact translation of viral proteins [38].

Conclusion:
We report the codon usage patterns of Sabia and Chapare viruses relative to the host codon usage pattern. Data shows a weak codon bias in Sabia and Chapare viruses to help in adaptation to the host. Mutation is affecting variation in codon patterns of viral sequences than hydropathicity and aromaticity. Thus, the data reported from this study help in understanding the mechanism of viral adaptation inside the host genome for further consideration in drug discovery.

References:
[1] Pellett PE *et al*. *Basics of virology* 2014 123:45 [PMID: 25015480].
[2] Shao J *et al*. *Pathogens* 2015 4:283 [PMID: 26011826].
[3] Hallam SJ *et al*. *Frontiers in Microbiology* 2018 9:1751 [PMID: 30123198].
[4] Radoshitzky R Sheli *et al*. *Archives of Virology* 2015 160:1851 [PMID: 25935216].
[5] Chiara F *et al*. *Journal of Molecular Biology* 2018 430:1839 [PMID: 29705070]
[6] Casals J *Yale J Bio Med.* 1975 48:15 [PMID: 168692].
[7] Brisse EM & Ly H *Front Immunol.* 2019 10:372[PMID: 30918506].
[8] Sarute N & Ross SR *Annual Review of Virology* 2017 4:141 [PMID: 28645238].
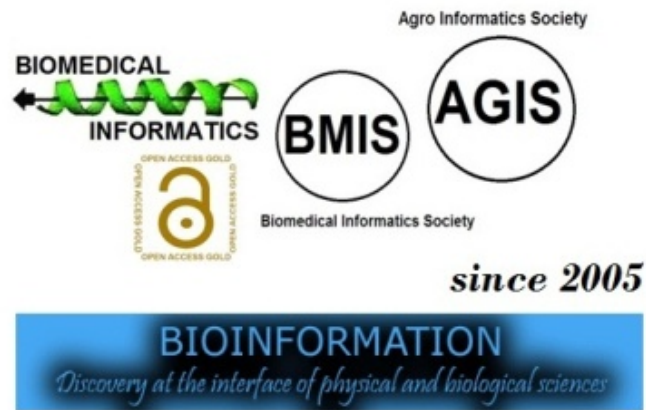[9] Gowen BB & Bray M. *Future Microbiology* 2011 6:1429 [PMID: 222122440].

©Biomedical Informatics (2021)

[10] Malta Mello de F *et al. Emerging Infectious Diseases* 2020 26:1332 [PMID: 32441627].

[11] Delgado S *et al. PLoS Pathogens* 2008 4:e1000047 [PMID: 18421377].

[12] Roy A *et al. Frontiers in Microbiology* 2017 8:1083[PMID: 28663742].

[13] Romero H *et al. Gene* 2003 317:141[PMID: 14604802].

[14] Gu W *et al. Biosystems* 2004 73:89[PMID: 15013221].

[15] Sharp PM *et al. Biochemistry Society Transactions* 1993 21:835 [PMID: 8132077].

[16] Shackelton L. A *et al J. Mol. Evol.*2006 62:551[PMID: 16557338].

[17] Roy A *et al. World Journal of Microbiology and Biotechnology* 2015

[18] 31:959[PMID: 25842224].

[19] Benson DA *et al. Nucleic Acids Research* 2017 45:D37 [PMID: 27899564].

[20] Pickett BE *et al. Nucleic Acids Research* 2012 40:D593 [PMID: 22006842].

[21] Puigbò P *et al. Biology Direct* 2008 3:38 [PMID: 18796141].

[22] http://www.molbiol.ox.ac.uk/cu

[23] Chen L *et al. Biochemical and Biophysical Research Communications* 2013

[24] 430:1344 [PMID: 23268345].

[25] Frey, Felix. (2017). SPSS (software). 10.1002/9781118901731.iecrm0237.

[26] Behura S K & Severson D W *PLoS One* 2012 7:e43111 [PMID: 22912801].

[27] Yi S *et al. Genomics* 2018 110:134[PMID: 28911975].

[28] Ramaiah A *et al. Infect Genet Evol.* 2017 51:74[PMID: 28315476].

[29] Puigbo P *et al. BMC Research Notes* 2010 3:87 [PMID: 20356391].

[30] Nasrullah I *et al. BMC Evolutionary Biology* 2015 15:174 [PMID: 26306510].

[31] Reis dos M *et al. Nucleic Acids Res.* 2004 32:5036 [PMID: 15448185].

[32] Duret L *Trends in Genetics* 2000 16:287[PMID: 10858656].

[33] Jiang W *et al. Biotechnology and Applied Biochemistry* 2017 64:218 [PMID: 27696508].

[34] Bahir I *et al. Molecular Systems Biology* 2009 5:311 [PMID: 19888206].

[35] Wright F *Gene* 1990 87:23[PMID:2110097]

[36] Jia X *et al. BMC Genomics* 2015 16:356[PMID: 25943559].

[37] Amicis De Fa & Marchetti S *Nucleic acids research* 2000 28:3339 [PMID: 10954603].

[38] Ohto U *et al. Nature* 2015 520:702 [PMID: 25686612].

[39] Li X *et al. International Journal of Molecular Sciences* 2016 17:1138 [PMID: 27428961].

[40] Reid CR *et al. Viruses* 2015 7:4385[PMID: 26287230].

Edited by P Kangueane

Citation: Malhotra & Kumar, Bioinformation 17(10): 891-898 (2021)

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.