CONSEQUENCES OF MUTATION, SELECTION AND PHYSICO-CHEMICAL PROPERTIES OF ENCODED PROTEINS ON SYNONYMOUS CODON USAGE IN ADENOVIRUSES *

SANDIP PAUL, SABYASACHI DAS and CHITRA DUTTA

Bioinformatics Centre, Indian Institute of Chemical Biology, Kolkata –700032, India

Trends in synonymous codon usage in adenoviruses have been examined through the multivariate statistical analysis on the annotated protein-coding regions of 22 adenoviral species, for which complete genome sequences are available. One of the major determinants of such trends is the G + C content at third codon positions of the genes, the average value of which varied from one viral genome to other depending on the overall mutational bias of the species. G_{3S} and C_{3S} interacted synergistically along the first principal axis of Correspondence analysis on the Relative Synonymous Codon Usage of adenoviral genes, but antagonistically along the second principal axis. Other major determinants of the trends are the natural selection, putatively operative at the level of translation and quite interestingly, hydropathy of the encoded proteins. The trends in codon usage, though characterized by distinct virus-specific mutational bias, do not exhibit any sign of host-specificity. Significant variations are observed in synonymous codon choice in structural and nonstructural genes of adenoviruses.

1. Introduction

Genomes of adenoviruses are characterized by linear, double-stranded DNA, with inverted terminal repeats (ITR) ranging from 36 to over 200 bp in length depending on the serotype.[1,2] Genes inherited by all existing adenoviruses from their common ancestor (Genus-common genes) are located centrally in the genome and are involved in replication and packaging of viral DNA as well as in the formation of virion. The other genes (Genus -specific genes) are captured in each lineage and mostly located near the genome termini.[2] These genes are generally involved in interactions with the host and probably contribute to the survival of viruses in respective biological niches. [2]

In recent years, the focus of adenovirus research has shifted from basic biology to adenovirus-based vector technologies.[3] Adenoviruses are often efficient at gene delivery to specific cell types. Genetically engineered, replication-deficient recombinant adenoviruses are gradually becoming popular as gene delivery vehicles for their high capacity to transfer therapeutic genes *in vivo*.[4] One of the crucial issues for development of promising vectors for gene therapy are transient, but high level of expression of delivered genes within the host. But development of an efficient gene expression system

* This work was supported by the Council of Scientific and Industrial Research, Government of India (Project No. CMM 0017) and Department of Biotechnology, Government of India (Grant Number BT/BI/04/055-2001).

needs a detailed knowledge of codon and nucleotide preferences in genomes concerned. Keeping this in mind, the present study attempts to analyze the nucleotide and codon usage patterns in all adenoviral genomes sequenced so far.

It is well established that synonymous codon usage in various organisms, particularly in unicellular ones, often reflect a balance between the genomic G+C-bias and translational selection.[5,6] The strength and direction of these selection forces vary at the intra- and inter-genomic level.[7] Various other factors like codon-anticodon interaction,[8] physical location of each gene on the chromosome,[9] replicational-transcriptional selection,[10-12] ecological niches [13] etc. may influence the biased usage of synonymous codons. In viruses, however, little is known about the extent and origin of the synonymous codon bias. In human immunodeficiency virus (HIV), codon usage bias is the result of strong preference of adenine base.[14] Codon bias due to uneven base composition has also been described in nucleopolyhedroviruses [15] and pneumoviruses.[16] In papillomaviruses, specific codon usage pattern linked with variation in A+T-content within the genomes may increase the replicational fitness in mammalian epithelial cells.[17] In human RNA viruses, mutational bias is not the only determining factor, translational selection may also have influence in shaping codon usage bias.[18] In the present report, through a multivariate analysis, attempt has been made to delineate the trends in codon and nucleotide selection in adenoviral genes and also to identify the selection forces governing such trends. Such information not only can offer an insight into the evolution of codon usage patterns along adenovirus lineages, but also may help in increasing the efficiency of gene delivery/expression systems.

2. Methods and Materials

2.1 Retrieval of sequences

The 22 available complete genome sequences of adenoviruses (listed in Table1) have been downloaded from NCBI GenBank (Version 145.0). To minimize the sampling error we have taken only those genes, which are greater than or equal to 150 bp. We have also eliminated the partial coding sequences and those sequences, which have internal termination codons. Finally 616 coding sequences were selected for analysis.

2.2 Sequence analysis

Relative synonymous codon usage (RSCU) was used to examine the synonymous codon usage variation among the genes without any confounding influence of amino acid composition.[5] To find out the extent of base compositional bias GC_{1+2} (G+C content at first and second codon positions), GC_{3S} and N_{3S} (the frequency of G+C and base N respectively at synonymous third codon positions) were calculated for each gene under study. To measure the general non-uniformity of synonymous codon usage, the effective number of codons (Nc) of each gene was calculated.[19] The GRAVY score, which indicates the mean hydropathy index of the encoded amino acid residues and hence, is an

estimate of overall hydrophobicity,[20] was computed for each gene product. The predictions of protein secondary structure were performed using GOR IV algorithm.[21]

Correspondence analysis (COA) on RSCU values was carried out using CODONW 1.4.2 to investigate the major trend in codon usage variation among genes. To see the extent of divergence in codon usage more precisely, a cluster analysis was carried out using simple D-squared statistic method. D-squared statistics is the sum of squares of the difference between codons of the two codon usage tables; i.e. D^2 = sum over all 64 codons of: $(\text{frequency}_{(\text{codon, Table 1})} - \text{frequency}_{(\text{codon, Table 2})})^2$. A matrix containing the D^2 value of each pair was used to produce clusters (dendrogram) by neighbor-joining method.[22] Linear regression analysis was used to find out the correlation between synonymous codon usage bias and various codon usage indices. To test the heterogeneity in codon usage, one-way ANOVA was performed using STATISTICA (Version 6.0).

3. Results and Discussion

3.1 Inter-and Intra-species variation in compositional constraints on codon usage

The codon usage bias in the coding regions of 22 completely sequenced adenoviruses of varying G+C content has been investigated (Table1). The average values of the effective numbers of codons (Nc) in different adenoviruses varied from 38.97 (in *Porcine adenovirus A*) to 54.67 (in *Canine adenovirus*). The average GC_{3S} values for individual genomes varied from 22.78 (OAdV-A) to 79.61 (PAdV- A). In addition, there are marked intra-genomic variations in Nc (standard deviation > 3.5, except for BAdV-B) and GC_{3S} values (standard deviation > 5%). These observations indicate that there is a significant heterogeneity in compositional bias as well as in the codon usage pattern within and among the members of *Adenoviridae*. When the Nc values of each adenovirus gene are plotted against the corresponding GC_{3S}, only a small number of points lie on the expected curve and a large number of points lie well below the expected curve (not shown), suggesting that some additional selection pressure other than the species-specific mutational bias, acts on codon usage in adenovirus genomes.

3.2 Virus-specific synonymous codon usage patterns with no sign of host-specificity

Figure 1 depicts the position of each virus on the plane defined by the first (horizontal) and second (vertical) principal axes generated by COA on RSCU values of genes. The first and second principal axes account for 30.35 % and 7.6 % of total variability. The first principal axis exhibits strong correlation with GC_{3S}-content of the genes for all four genera of adenoviruses i.e. for atadenovirus, aviadenovirus, mastadenovirus and siadenovirus (Table 2). The viruses having highest GC_{3S} levels in their coding sequences display the most positive values along that axis (Fig. 1). When G_{3S} and C_{3S} are considered separately, the correlation coefficient exhibits by the positions of genes along the first axis with C_{3S} is significantly larger than that with G_{3S} (Table 2), indicating that

the contribution of C_{3S} to the inter-species variation in overall GC_{3S}-content is greater than that of G_{3S}.

Table 1. Average codon usage bias (measured by effective number of codons) and base composition in 22 completely sequenced adenoviruses under study.

Virus	Abbreviation	Accession	GC %	Nc	A_{3S}	T_{3S}	G_{3S}	C_{3S}
ATADENOVIRUS								
Bovine adenovirus D	BAdV-D	NC_002685	35.2	44.46	32.93	39.05	16.53	11.49
Duck adenovirus A	DAdV-A	NC_001813	43.0	53.46	25.90	34.75	21.42	17.92
Ovine adenovirus D	OAdV-D	NC_004037	33.6	42.56	32.89	41.38	14.93	10.80
AVIADENOVIRUS								
Fowl adenovirus A	FAdV-A	NC_001720	54.3	52.36	16.58	21.66	28.05	33.71
Fowl adenovirus D	FAdV-D	NC_000899	53.8	51.01	17.11	20.43	28.05	34.41
MASTADENOVIRUS								
Bovine adenovirus A	BAdV-A	NC_006324	48.8	54.27	22.94	27.04	25.84	24.18
Bovine adenovirus B	BAdV-B	NC_001876	54.0	51.67	15.68	23.53	28.42	32.37
Canine adenovirus	CAdV	NC_001734	47.0	54.67	22.38	29.48	23.38	24.76
Human adenovirus A	HAdV-A	NC_001460	46.5	54.15	25.32	30.59	21.85	22.24
Human adenovirus B	HAdV-B	NC_004001	48.9	51.88	22.95	28.78	22.41	25.86
Human adenovirus C	HAdV-C	NC_001405	55.2	47.21	15.72	21.83	30.46	31.99
Human adenovirus D	HAdV-D	NC_002067	56.6	46.40	15.53	19.73	29.67	35.07
Human adenovirus E	HAdV-E	NC_003266	57.7	44.47	12.39	18.17	31.10	38.34
Human adenovirus F	HAdV-F	NC_001454	51.2	51.99	18.73	27.36	25.8	28.11
Murine adenovirus A	MAdV-A	NC_000942	47.8	53.09	22.35	31.36	24.28	22.01
Ovine adenovirus A	OAdV-A	NC_002513	43.6	49.73	26.94	31.55	20.63	20.88
Porcine adenovirus A	PAdV-A	NC_005869	63.8	38.97	7.63	12.66	35.99	43.72
Porcine adenovirus C	PAdV-C	NC_002702	50.5	50.64	18.48	23.55	27.02	30.95
Simian adenovirus A	SAdV-A	NC_006144	55.3	47.36	15.81	22.62	29.20	32.37
Tree shrew adenovirus	TSAdV	NC_004453	50.0	50.42	19.95	26.87	25.87	27.31
SIADENOVIRUS								
Frog adenovirus	FrAdV	NC_002501	37.9	44.81	33.98	35.10	16.76	14.16
Turkey adenovirus A	TAdV-A	NC_001958	34.9	44.09	30.53	41.34	16.49	11.64

The separation of one viral genome from another is found to be significant by one-way analysis of variance (ANOVA) on the first principle axis ($F_{21, 594} = 57.805$, $p < 10^{-7}$), as that axis explains the major variation in codon usage. This result indicates that some virus-specific selection pressure might have influenced synonymous codon usage in different adenoviruses. However, no sign of host-specificity can be observed in the trends in codon usages. Viruses infecting the same host appear, in most cases, at distinct positions along axis1 and/or axis2 generated by COA on RSCU values of genes (Fig. 1). For example, the viruses infecting the bovine host (BAdV-A, BAdV-B and BAdV-D) are placed in three different positions significant distances apart from one another. Therefore, it seems that synonymous codon usage patterns in adenoviruses do not follow, in general, any host specific trend.

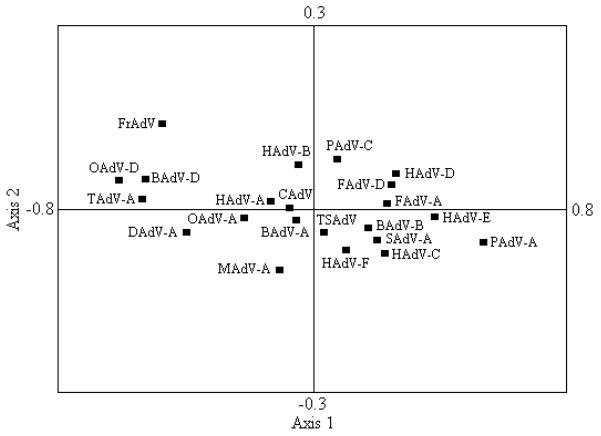

Figure 1: Positions of adenoviruses on the plane defined by first and second principle axes generated from Correspondence analysis of Relative Synonymous Codon Usage (RSCU) of corresponding genes. GC-poor species are on the left, while GC-rich on the right.

The members of different genera of adenoviruses (atadenovirus, aviadenovirus, mastadenovirus, siadenovirus) exhibit non-uniform distribution along the axis 1 in Fig. 1. The position of atadenoviruses and siadenoviruses are at the extreme left of axis1, whereas aviadenoviruses exhibit the positive values along this axis. The separation of these four genera on the basis the variation in codon usage pattern explained by first axis of COA on RSCU values is statistically significant ($F_{3, 612} = 161.35$, $p < 10^{-7}$). This observation indicates that so far as the synonymous codon usage are concerned, members of atadenovirus and siadenovirus are more close to one another, while those in aviadenoviruses are far away from these two genera. The members of mastadenovirus are distributed over a large region, suggesting that this group follows more heterogeneous patterns in synonymous codon usage than the other adenoviruses. These observations are in accordance with the cluster analysis on the extent of divergence in codon usage, which yields two major clusters - the adenoviruses having relatively higher GC-content, e. g., aviadenoviruses are branched together in the upper cluster and those with relatively AT- rich genomes, such as atadenoviruses and siadenoviruses, appear in the lower cluster (Fig. 2). Members of mastadenovirus, having widely varying GC-content, are dispersed through both the clusters. It is worth mentioning that the distribution of different adenoviral genomes along the axis1 (Fig. 1) and their segregation in two clusters (Fig. 2) are consistent with their genomic G+C-content and also with their phylogenetic distribution, as determined by Davison et al (2003).[2]

Table 2. Non-parametric tests of association between the first two axes of COA on RSCU and multiple synonymous base usage parameters and hydropathy of encoded proteins

Parameters	Atadenovirus		Aviadenovirus		Mastadenovirus		Siadenovirus	
	Axis 1	Axis2	Axis 1	Axis2	Axis 1	Axis2	Axis 1	Axis2
A_{3S}	-0.49***	0.83***	-0.42**	0.46***	-0.77***	0.39***	-0.35*	0.66***
T_{3S}	-0.52***	-0.48***	-0.86***	-0.50***	-0.90***	-0.19**	-0.57***	-0.61***
G_{3S}	0.55***	-0.76***	0.04^{NS}	-0.81***	0.54***	-0.67***	0.51**	-0.60***
C_{3S}	0.81***	0.28*	0.89***	0.74***	0.89***	0.43***	0.82***	0.47**
GC_{3S}	0.95***	-0.35**	0.92***	0.17^{NS}	0.97***	-0.06^{NS}	0.90***	-0.05^{NS}
Gravy	0.14^{NS}	-0.51***	-0.45**	-0.09^{NS}	-0.41***	-0.03^{NS}	0.02^{NS}	-0.25*

Notable significant relationships are marked by *** $P < 0.0001$;** $P < 0.001$;* $P < 0.01$;^{NS} Nonsignificant.

The correlation coefficient between the second axis and GC_{3S} is relatively small, as compared to that between the axis1 and GC_{3S} (Table 2). But it is worth mentioning that the axis2 exhibits strong negative correlation with G_{3S} and positive correlation with C_{3S} for all four genera of adenoviruses (Table 2). These observations indicate that G_{3S} and C_{3S} interact synergistically in the first principal axis resulting in the increase of GC_{3S} content, but antagonistically in the second principal axis so that increase in the frequency of C_{3S} is accompanied by a decrease in G_{3S} and vice-versa.

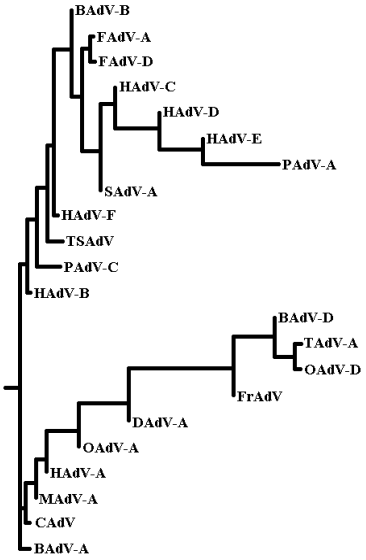

Figure 2: A dendrogram representing the extent of divergence in synonymous codon usage of 22 completely sequenced adenoviruses constructed by neighbor-joining method.

This antagonistic behavior might be due to the fact that after GC_{3S} reach some saturation value, frequency of one (G_{3S} or C_{3S}) of them can increase further only at the expense of the other one (C_{3S} or G_{3S}), i.e., occurrence of one base excludes the other in order to maintain the overall G+C-content at synonymous positions under some threshold value.

3.3 Differential codon usage in structural and nonstructural genes

With a view to examine whether the presence of any selection pressure(s) is responsible for intra-genome heterogeneity in codon usage patterns in adenoviruses, we have compared the average RSCU values of structural (i.e. major core protein, minor core protein, hexon and hexon-associated protein etc.) and nonstructural (i.e. DNA polymerase, transcription activators etc.) genes of four different adenoviral genera separately. There are several codons (mostly G- or T-ending), usages of which are significantly higher among the structural genes. On the other hand, several codons (C- or A-ending) are over represented in nonstructural genes as compared to the codon usage in structural genes (data not shown). This indicates that the selection force resulting in the differential codon usage patterns in structural and non-structural genes might not be the simple G+C-bias. As the structural genes in viruses are generally highly expressed than the nonstructural genes,[23] a natural selection, putatively operating at the level of translation might also be responsible for differential usage of synonymous codons in

structural and non-structural genes of adenoviruses. However, in adenoviruses, no significant correlation is found between codon usage bias (as measured by Nc values of genes) and the gene length.

3.4 Correlation between synonymous base compositions and hydropathy of encoded proteins

An important finding derived from our study on adenoviral genomes is that for each group of adenoviruses, one of the principal axes generated by COA on RSCU values of the genes exhibits significant correlation with the hydropathy of the encoded proteins (as determined by the Gravy Score of the gene-products) (Table 2). With a view to find out the reason behind this apparently unexpected association between nucleotide usage in third codon position and protein hydropathy, we have calculated the correlations between hydrophobicity and synonymous base usage in four adenovirus genera separately. It is found that the hydrophobicity exhibits highest positive correlations with G_{3S} in atadenoviruses ($r = 0.43$, $p < 0.001$) and siadenoviruses ($r = 0.36$, $p < 0.001$), whereas with T_{3S} in aviadenoviruses ($r = 0.37$, $p < 0.001$) and mastadenoviruses ($r = 0.54$, $p < 0.001$) (Fig. 3).

Usage of individual nucleotides at synonymous sites of genes in members of different genera of adenoviruses reveals a mixed and indefinite nature of the correlation with hydrophobicity of respective gene-products (Table 3). In general (with a few exceptions), the hydropathy levels of encoded proteins shows positive correlation with both G_{3S} and T_{3S} in atadenoviruses, with only T_{3S} in aviadenoviruses and mastadenoviruses, and with only G_{3S} in siadenoviruses.Earlier it is found that as compared to hydrophilic proteins, there is an increase in usage of the G-ending codons and decrease in that of the C-ending codons in hydrophobic proteins of *Mycobacterium tuberculosis* and *Mycobacterium leprae*.[24] Existence of significant correlation between hydrophobicity of encoded proteins and the base composition of third codon positions have also been reported in some other prokaryotes and several eukaryotes.[25] However, there is no report of such correlation in any of the viral genomes studied so far. To our knowledge, this is for the first time; a correlation has been demonstrated between the synonymous codon usage in genes and this physico-chemical property of corresponding gene-products in a group of viral genomes.

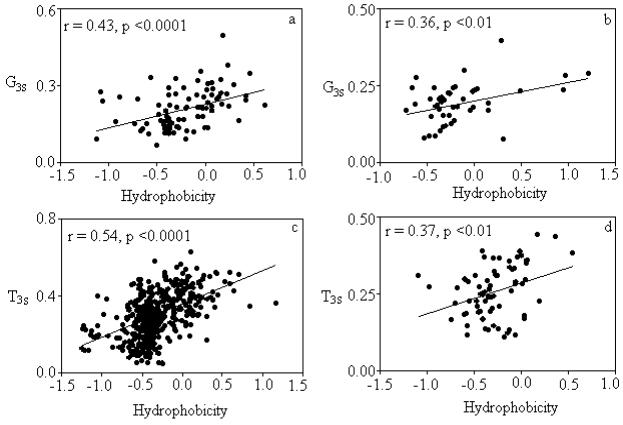

Figure 3: Hydropathy values (Gravy scores) of the encoded proteins plotted against (a) G_{3S} in atadenoviruses, (b) G_{3S} in siadenoviruses, (c) T_{3S} in mastadenoviruses and (d) T_{3S} in aviadenoviruses.

Table 3. Correlations between the protein hydropathy and synonymous base compositions in adenoviruses

Genera	Virus	Correlation coefficients (r)			
		A3S[a]	T3S[b]	G3S[c]	C3S[d]
Atadenovirus	BAdV-D	-0.73**	0.51**	0.61**	-0.43*
	DAdV-A	-0.44*	0.09	0.25	0.15
	OAdV-D	-0.54**	0.51**	0.41*	-0.33
Aviadenovirus	FAdV-A	-0.07	0.28	-0.02	-0.27
	FAdV-D	0.22	0.43*	-0.01	-0.38*
Mastadenovirus	BAdV-A	-0.06	0.58**	-0.41*	-0.32
	BAdV-B	0.19	0.72**	-0.24	-0.43*
	CAdV	0.01	0.60**	0.02	-0.59**
	HAdV-A	-0.04	0.54**	-0.05	-0.57**
	HAdV-B	0.34*	0.36*	-0.39*	-0.22
	HAdV-C	-0.01	0.63**	-0.20	-0.49**
	HAdV-D	0.43*	0.44*	-0.42*	-0.31
	HAdV-E	0.32	0.58**	-0.48**	-0.41*
	HAdV-F	0.29	0.72**	-0.39*	-0.53**
	MAdV-A	-0.44*	0.41*	0.08	-0.11
	OAdV-A	0.47*	0.60**	-0.43*	-0.58**
	PAdV-A	0.13	0.41*	0.09	-0.31
	PAdV-C	0.25	0.58**	-0.30	-0.47*
	SAdV-A	0.39*	0.65**	-0.42*	-0.56**
	TSAdV	0.37*	0.76**	-0.31	-0.66*
Siadenovirus	FrAdV	0.09	-0.06	0.12	-0.48*
	TAdV-A	-0.33	-0.24	0.58**	0.05

Significant relationships are marked by ** $P < 0.01$; * $P < 0.05$

Therefore, the present analysis indicates that the selection of nucleotide at synonymous sites in adenoviral genes might affect or be affected by the hydropathy levels of the encoded products. However, the cause and effect relation of this correlation is not clear. It is known that the hydrophobicity of amino acid residues plays an important role in protein folding.[25] It has also been reported that the codons overrepresented in alpha-helix are underrepresented in beta-sheet and vice versa and this discrepancy may be related to the particular translation kinetics necessary to ensure the proper folding of nascent peptide.[24] With a view to predict the plausible biological origin of the correlation between synonymous codon usage and protein hydropathy, we have, therefore, predicted the secondary structure of five most hydrophobic proteins (i.e., gene-products with highest Gravy scores) and five most hydrophilic proteins (i.e., gene-products with lowest Gravy scores) from each adenovirus species. It is found that the regions of the proteins with high propensity of formation of alpha-helices are significantly over represented (t-test, p-value$<10^{-5}$) in proteins with high hydrophilicity (i.e., low Gravy Score), whereas the regions predicted to be beta-sheet were found in significantly higher frequencies (t-test, p-value$<10^{-5}$) in proteins with high hydrophobicity (i.e., high Gravy Score). Therefore, the differential synonymous base

usage by hydrophobic and hydrophilic proteins in adenoviruses may have its origin in their propensities for secondary structure formation. Further studies are required to get the deeper insights on the biological factors underlying this relationship in viruses.

In summary, the trends in synonymous codon usage in adenoviruses is found to be governed by several factors - the virus-specific directional mutational bias, natural selection putatively operating at the level of translation and more interestingly, hydrophobicity of the gene-products. Apparently, the trends in synonymous codon selection do not exhibit any host-specificity. No correlation is found between codon usage bias and gene-length. Furthermore, the antagonistic behavior of G_{3S} and C_{3S} along the second major axis of COA on RSCU values of adenoviral genes suggests the existence of a constraint on the extent of GC-bias in third codon position of any specific adenoviral genome. Such findings on trends in synonymous codon usage in adenoviruses might provide not only the valuable information for better understanding of the evolution of adenoviral genomes, but also provide clues to development of an efficient gene delivery/expression systems based on adenoviral vectors.

References

1. R. N. de Jong, P. C. van der Vliet and A. B. Brenkman. Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr. Top. Microbiol. Immunol.,* 272;187–211, 2003.
2. A. J. Davison, M. Benko and B. Harrach. Genetic content and evolution of adenoviruses. *J. Gen. Virol.*, 84;2895-2908, 2003.
3. J. J. Rux and R. M. Burnett. Adenovirus structure. *Hum. Gene. Ther.*, 15;1167-1176, 2004.
4. C. M. Lai, Y. K. Lai, and P. E. Rakoczy. Adenovirus and adeno-associated virus vectors. *DNA Cell Biol.,* 21;895-913, 2002.
5. P. M. Sharp and W. H. Li. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.*, 24;28-38, 1986.
6. A. Pan, C. Dutta, and J. DAS. Codon usage in highly expressed genes of *Haemophilus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias. *Gene,* 215;405-413, 1998.
7. S. Das, A. Pan, S. Paul, and C. Dutta. Comparative Analyses of Codon and Amino Acid Usage in Symbiotic Island and Core Genome in Nitrogen-Fixing Symbiotic Bacterium Bradyrhizobium japonicum. *J. Biomol. Struct. Dyn.,* In Press., 2005b.
8. H. Grosjean and W. Fiers. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*, 18;199-209, 1982.
9. A. R. Kerr, J. F. Peden, and P. M. Sharp. Systematic base composition variation around the genome of Mycoplasma genitalium, but not *Mycoplasma pneumoniae*. *Mol. Microbiol.,* 125;1177-1179, 1997.
10. J.O. McInerney. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. USA,* 95;10698-10703, 1998.

11. H. Romero, A. Zavala, H. Musto. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.,* 28;2084-2090, 2000.
12. S. Das, S. Paul, S. Chatterjee and C. Dutta. Codon and Amino Acid Usage in Two Major Human Pathogens of Genus *Bartonella* – Optimization Between Replication-Transcriptional Selection, Translational Control and Cost Minimization. *DNA Res.,* 12;91–102, 2005a.
13. G. A. Singer and D. A. Hickey. Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene,* 317;39-47, 2003.
14. F. J. van Hemert, and B. Berkhout. The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. *J. Mol. Evol.,* 41;132-140, 1995.
15. D. B. Levin and B. Whittome. Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.,* 81;2313-2325, 2000.
16. C. R. Pringle and A. J. Easton. Monopartite negative strand RNA genomes. *Seminars in Virology,* 8;49-57, 1997.
17. K. N. Zhao, W. J. Liu and I. H. Frazer. Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Res.,* 98;95-104, 2003.
18. G. M. Jenkins and E. C. Holmes. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.,* 92;1-7, 2003.
19. F. Wright. The 'effective number of codons' used in a gene. *Gene,* 87;23-29, 1990.
20. J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.,* 157;105-132, 1982.
21. J. Garnier, J. F. Gibrat and B. Robson. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.,* 266;540-553, 1996.
22. N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.,* 4;406-425, 1987.
23. W. Gu, T. Zhou, J. Ma, X. Sun and Z. Lu. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res.,* 101;155-161, 2004.
24. A. B. de Miranda, F. Alvarez-Valin, K. Jabbari, W. M. Degrave and G. Bernardi. Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *J. Mol. Evol.,* 50;45-55, 2000.
25. G. D'Onofrio, K. Jabbari, H. Musto and G. Bernardi. The correlation of protein hydropathy with the base composition of coding sequences. *Gene,* 238;3-14, 1999.