Original article

A complete analysis of Relative Synonymous Codon Usage in HVRs1-4 region in adenovirus genome

J.S. Niczyporuk[1]

[1]Department of Poultry Diseases, National Veterinary Research Institute,
Partyzantow 57, 24-100 Pulawy, Poland

Abstract

Recent outbreaks of adenovirus (FAdV) infections in poultry flocks have been determined in many countries in Europe, Asia and Australia connected with economic consequences, and loses in poultry production. To better understand the evolution and transmission of FAdV viruses, detailed codon usage analysis was performed for 137 recently obtained FAdV strains. A high effective number of codons, and an indication the presence of low codon usage were determined. The presence of mutations, and their influence on codon usage was confirmed by a correlation between nucleotide compositions at the 3rd codon positions, HVRs1-4, and ENCs. This presence indicate some influence of natural selection, and antigenic properties of examined FAdV strains.

Key words: fowl adenoviruses, genome, codon usage, HVRs

Introduction

Adenoviruses (FAdVs) are wildely distributed worldwide in the poultry industry. Their pathogenic role is not always clear. They can cause latent infection or several diseases. Data concerning the adenovirus genome, structure, function of selected proteins, genome organization and replication are based on human adenovirus strains (Rux et al. 2003, Benko et al. 2005). In Poland the presence of five species with eight types of adenovirus strains: FAdV-1/A, 2/D, 4/C, 5/B, 7/E, 8a/E, 8b/E, and 11/D was described recently (Niczyporuk 2014). The biggest group was formed by the strains classified as type FAdV-2/11/species D, and the smallest group by the type FAdV7/E (Niczyporuk 2014). Moreover, a relationship between types/species and the internal organs from which the strains were isolated was indicated. It was demonstrated that types/species FAdV-2/11/D were most commonly isolated from the liver and intestines while type FAdV-4/C was represented less frequently (Niczyporuk 2014). Avian adenovirus genome size is 44-45 kb depends from types/species representation. For comparison, mastadenoviruses have a genome size of 31-36 kb. The genome of adenoviruses belongs to different species and has different GC pair content. The value in avian adenoviruses is between 53.8% and 66.9% and in mastadenoviruses is 43.6% - 63.9% (Benko et al. 2005, Harrach and Kajan

Correspondence to: J.S. Niczyporuk, e: mail: jowita.niczyporuk@piwet.pulawy.pl

2011). Different sizes of the genome have an influence on gene organization, are specific dependent, and have from 23 to 46 genes which code several proteins (Chiocca et al. 1996). Number of genes and their functions is different among types/species strains; however their basic genome organisation is characteristic for the *Adenoviridae* family, especially in the central part. The genome contains over a dozen transcriptional units, which code between 1 and 8 sequences with open reading frames (ORF). Thanks to alternative splicing, between 50 and 70 different adenovirus proteins can be created. Till now, not all the genes and their functions are known (Corredor et al. 2008, Harrach 2008). The avian adenovirus genome has unique transcriptional units, so it has parts which are not well characterised yet. Adenoviruses have the ability for interferon induction (Toth et al. 1987) what was demonstrated in experimentally infected animals (Berencsi et al. 1993) and cell cultures (Beladi et al. 1979). The adenovirus genome can be also used as system for gene expression analysis (Kay et al. 1995, Russell 2000, Xu et al. 2007 Knowles 2011,). Natural selection and mutation pressure have an influence on the synonymous codon used during protein translation (Choludhury et al. 2017).

Relative synonymous codon usage (RSCU) is a simple measure of the heterogeneity in the usage pattern of synonymous codons. The value was calculated and performed by others (Sharp and Li 1986, Hassan et al. 2009, Fox and Erill 2010, Gu et al. 2012, Cholundhury et al. 2017, Halder et al. 2017). RSCU value >1 means, that the codon is more frequently used than expected, while the RSCU value <1 means that the codon is less frequently used than expected (Choludhury et al. 2017). The effect of mutational pressure on codon usage was confirmed by correlation between nucleotide compositions at the 3rd codon position (Singh and Tyagi 2017). Sharp and Li, 1986 also defined the RSCU value as the ratio of observed frequency of a specific codon to the frequency expected, and the effect of mutational pressure was assessed by correlations of the 3rd nucleotide position in the codon. The presence of RSCU values for type FAdV-1/A, and FAdV-7/E was reported by Niczyporuk, 2017.

Materials and Methods

Sequences of adenovirus field strains. Sequences of the Loop L1 region of the hexon gene covered HVRs1-4 from 137 adenovirus field strains designated as different types/species FAdV-2/11, 4, 5, 7, 8a, and 8b/FAdV-A-E were selected for sequence analysis.

Sequence of virus reference strain. The full-length sequence of reference strains, belonging to the types/

species FAdV-2/11/D, 4/C, 5/B, 7/E, 8a/E, and 8b/E derived from the GenBank database (NCBI) were used for nucleotide and amino acid sequence comparisons.

Strain isolation. The DNA of 137 FAdV strains was extracted from CEFs by using a QIAamp mini-kit (Qiagen, Germany) according to the manufacturer's instructions. The negative DNA controls were extracted from non-infected CEFs. The DNA was stored at -20°C for the next step of the study, as the template for sequencing. Extracted FAdVs DNA was tested by PCR to confirm the absence of other pathogens such as avian reovirus (ARV), chicken anemia virus (CAV), infectious bursa disease virus (IBDV), and Marek's disease (MD).

Sequencing and phylogenetic analysis. After amplification, 137 PCR products were purified using NucleoSpin Extract II (Marcherey-Nagel, France), and prepared for sequencing. The sequencing was performed by Genomed (Warsaw, Poland) using a GS FLX/Titanium sequencer (Roche, Switzerland). The phylogenetic analysis was performed by the alignment of nucleotide sequences of Loop L1 hexon gene originating from 137 field strains and designated according to the general obligatory classifications. An FAdV reference sequence was taken from the GenBank database. A phylogenetic tree was generated by the Neighbour-Joining method with the use of a p-distance model (on 1,000 bootstrapped data sets). The analyses were performed using: MEGA7, Geneious6 and BLAST. On the basis of these analyses, the relation between the examined strains were determined.

Nucleotide sequence analysis. The alignment of 137 FAdVs consensus sequences was prepared by aligning DNA sequences of strains classified to one types/species with MEGA7 and Geneious6.2 software. On this way seven consensus sequences representing seven adequate types/species were created. During the confirmation process for the correctness of the created consensus sequences, analysis of similarity with the reference sequence was conducted.

Amino acid sequence analysis. Obtained nucleotide consensus sequences were translated on adequate amino acid sequences. The analysis was conducted with Geneious6.2 software. Seven amino acid consensus sequences were created, and differences between type/species were determined.

HVRs amino acid sequence analysis. The presence of similarity and differences in amino acid HVRs-1-4 sequences was designated using Geneious6.2 software.

Codon usage analysis. In order to evaluate potential presence of codon preferences, relative synonymous codon usage and nucleotide codon composition in the examined region of the Loop L1 hexon gene was performed using Geneious6 software.
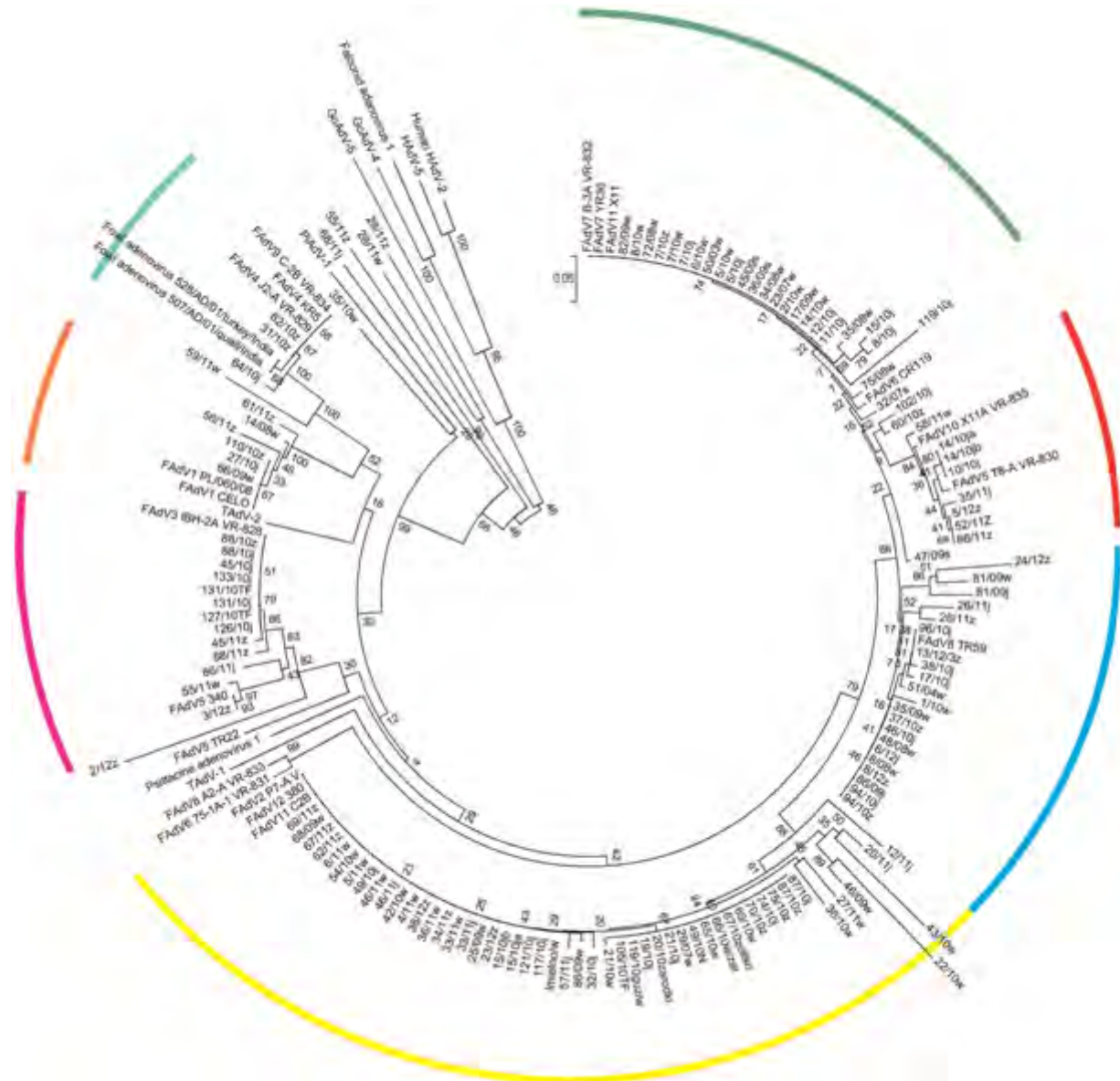
Fig. 1. Phylogenetic tree of 137 fowl adenovirusa (FAdVs) strains isolated in Poland. The phylogenetic tree have been root by the sequence derived from database GenBank: human adenovirus seguences (HAdV-2 i HAdV-5), Falconid adenowirus 1, PiAdV-1), goose adenovirus (GoAdV-4-5), (Psittacine adenowirus 1, PsAdV), and turkey adenovirus (TAdV-1). Colours indicate separate FAdVs types/species: green – FAdV-7, red – FAdV-8b, blue – FAdV-8a, yellow – FAdV-2/11 (FAdV-D), pink – FAdV-5, orange – FAdV-1, and deep blue – FAdV-4. The phylogenetic tree was generated by the Neighbour-Joining method with use of a p-distance model (on 1000 bootstraped data sets) based on nucleotide sequences of the hexon gene - L1 Loop HVRs-1-4 of fowl adenoviruses and reference strain sequence.

Results

Compositional properties. The nucleotide compositions (A, T, G, and C) and their location on the 1st, 2nd, and 3rd position in the codon in the coding sequence of the Loop L1 hexon gene, and the presence of (CG) contents were determined using Geneious6.2 software.

137 strains were divided into types/species: The branch for FAdV-1/A was formed by 6 strains. The branch for FAdV-2/11/D was created by 51 strains,

the branch for FAdV-4/C was created by 4 strains, the branch for FAdV-5/B was formed by 13 strains, the branch for FAdV-7/E by 22 strains, the branch for FAdV-8a by 21 strains, and the branch for FAdV-8b by 10 strains (Fig. 1). From these strains seven consensus sequences characteristic for type/species were created (Fig. 2). Almost no sequence variability was seen in separate types/species. Moreover, the translation of these sequences was conducted and seven amino acid consensus sequences were created (Fig. 3). For this analysis the sequence of reference strain FAdV-A/1(AF339914) was obtained. On this basis we

Fig. 2. Alignment of partial adenovirus hexon gene consensus nucleotide sequences. Graph presents nucleotide sequence. Over the examined sequences the graph presents the percentage of the variability for each nucleotide position from red, the most variable, to the green, the least variable region.

Fig. 3. Alignment of partial adenovirus hexon gene amino acid consensus sequences. The color of the background of amino acids indicates differences in relation to the adequate of amino acid sequence of reference strain FAdV-1/A. Red colour frames denote functionally conservative amino acids. Arrows over reference sequence indicate beginnings and ends of highly variable regions (HVR1-4). Over the examined sequences the graph presents the percentage of the variability for each nucleotide position from red, the most variable, to the green, least variable, region.
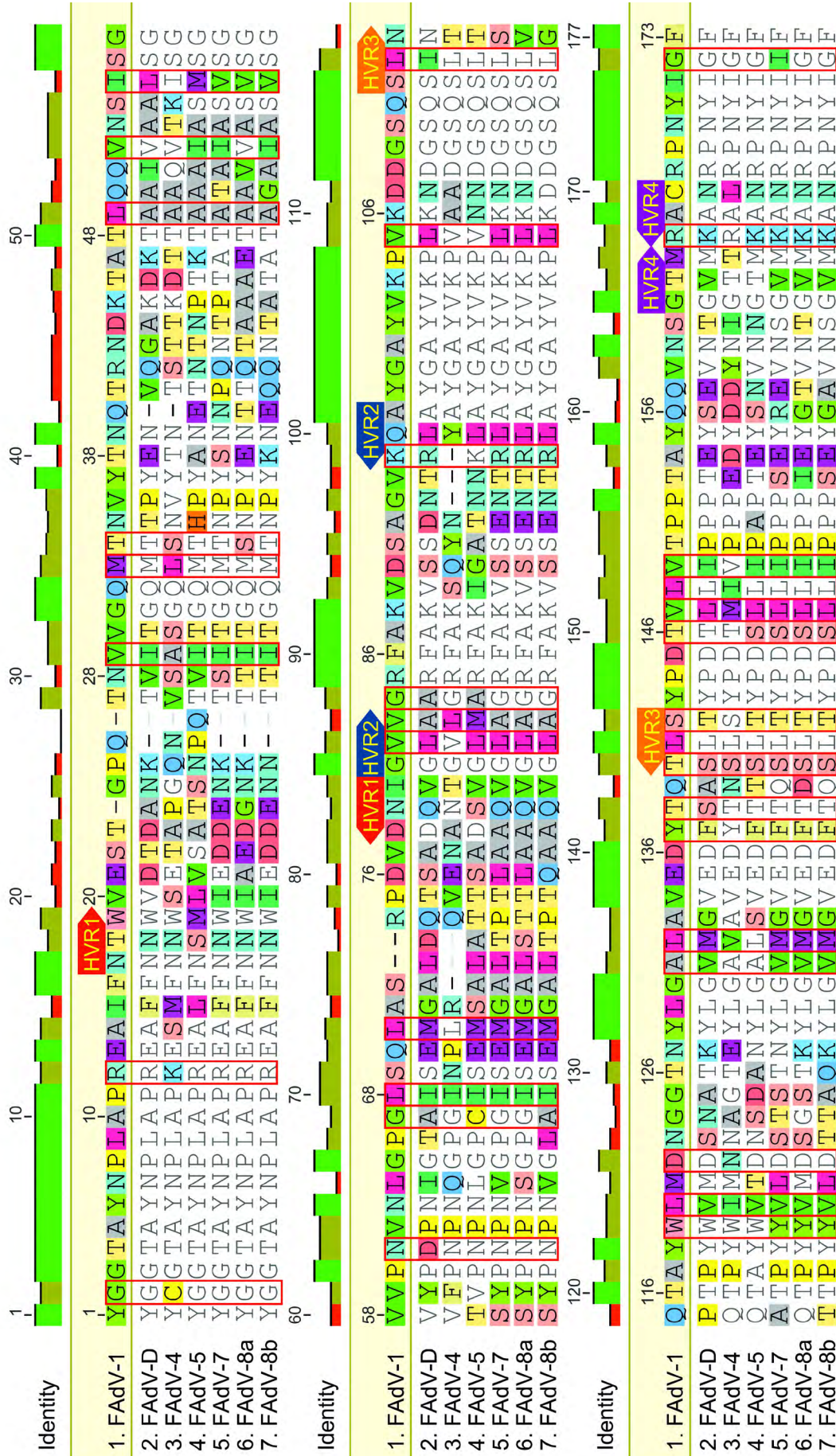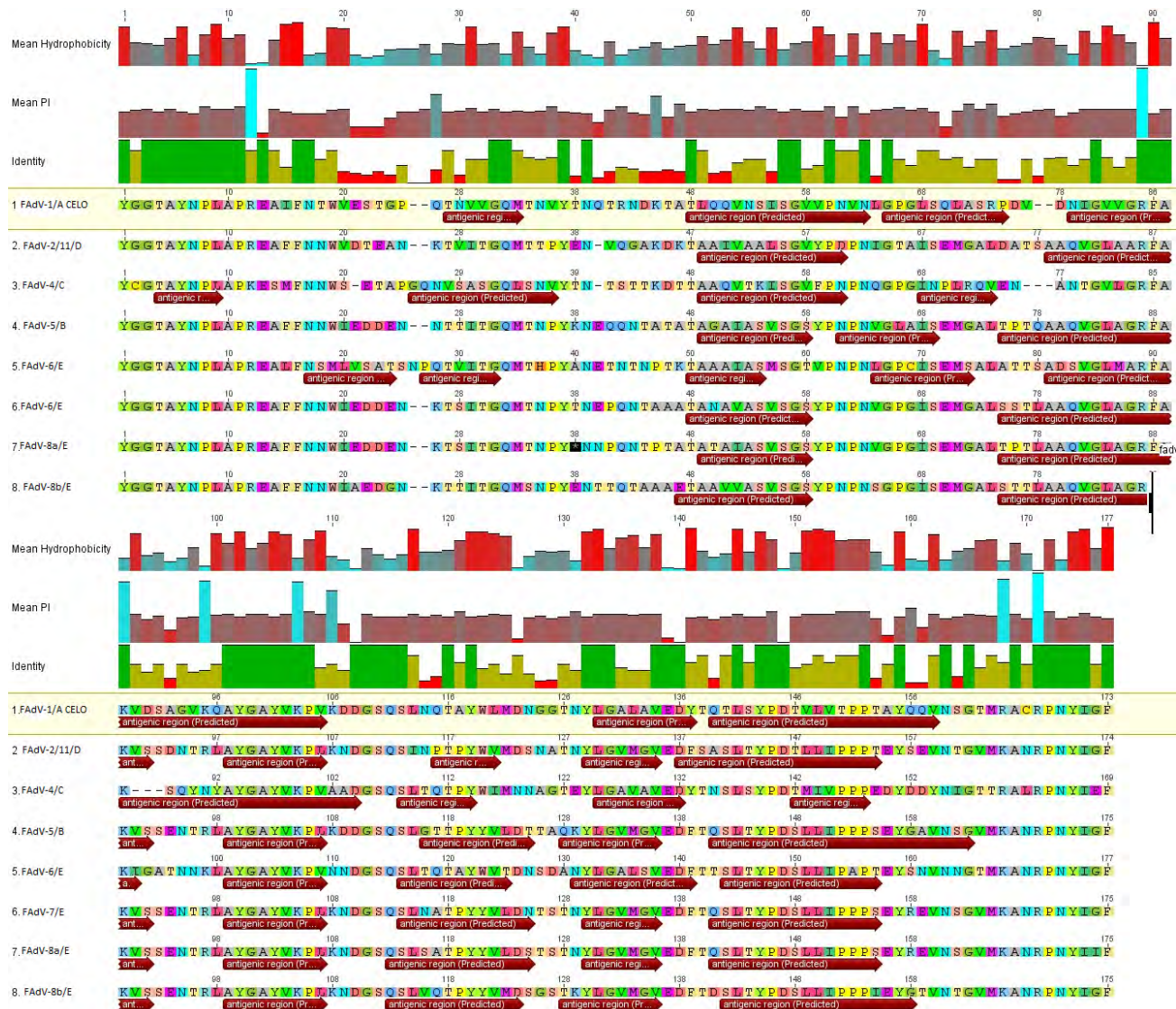
Fig. 4. Alignment of consensus amino acid sequences 1-177 aa of the examined adenovirus strains. Red arrows indicate the presence of predicted antigenic regions on the basics of amino acid sequence.

can indicate the differences and similarities between examined regions of the Loop L1 HVR-1-4 hexon gene of FAdVs strains.

The next step of the study was the designation if there was a codon preference in the examined region of HVR-1-4. For this reason the RSCU was calculated according to our own and reference strains. The analysis was conducted on 137 adenovirus strain sequences for which seven consensus sequences were created, and for the whole genome of the reference strain FAdV-A/1(AF339914) was determined in order to show the differences in codon preferences and antigenic regions (Fig. 4).

Seven consensus nucleotide (Fig. 2) and amino acid (Fig. 3) sequences of all examined strains demonstrated that from six codons which code Leu (L) for which optimised codon is prefered - codons, composed of three

nucleotide bases which specify amino acids during protein assembling CUG, (RSCU 2.53), and CUC (RSCU 1.42) respectively. Among four codons for Val (V), codon GUG (RSCU1.62), and GUC (RSCU 1.43) are preferred, optimised codons codes for Ala (A) are GCC (RSCU 1.76), and, in the case of codes for Ser (S) AGC (RSCU 2.04) and AGU (RSCU 0.73) are preferred. The presence of similarity and differences on HVRs-1-4 regions have been indicated by red arrows indicating preferred codons. Amino acids are in red columns (Fig. 3).

The sequence of codon analysis of the HVRs-1-4 region of the hexon gene for types/species indicated the differences in codon preference. For example, in most cases, types/species preferred codon CUG codes for Leu (L); however, in the case of strains from types//species FAdV-2/11/D and FAdV-4/C, codon CUC was

FAdV-1/A

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	0	0	UCU(S)	2	1.5	UAU(Y)	3	0.55	UGU(C)	0	0
UUC(F)	3	2	UCC(S)	2	1.5	UAC(Y)	8	1.45	UGC(C)	1	2
UUA(L)	0	0	UCA(S)	1	0.75	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	0	0	UCG(S)	0	0	UAG(*)	0	0	UGG(W)	2	1
CUU(L)	2	1.09	CCU(P)	1	0.4	CAU(H)	0	0	CGU(R)	1	1
CUC(L)	2	1.09	CCC(P)	7	2.8	CAC(H)	0	0	CGC(R)	0	0
CUA(L)	1	0.55	CCA(P)	1	0.4	CAA(Q)	2	0.33	CGA(R)	1	1
CUG(L)	6	3.27	CCG(P)	1	0.4	CAG(Q)	10	1.67	CGG(R)	1	1
AUU(I)	2	1.2	ACU(T)	1	0.25	AAU(N)	5	0.67	AGU(S)	2	1.5
AUC(I)	3	1.8	ACC(T)	11	2.75	AAC(N)	10	1.33	AGC(S)	1	0.75
AUA(I)	0	0	ACA(T)	1	0.25	AAA(K)	0	0	AGA(R)	1	1
AUG(M)	3	1	ACG(T)	3	0.75	AAG(K)	5	2	AGG(R)	2	2
GUU(V)	1	0.21	GCU(A)	4	1.07	GAU(D)	2	0.5	GGU(G)	3	0.71
GUC(V)	4	0.84	GCC(A)	5	1.33	GAC(D)	6	1.5	GGC(G)	6	1.41
GUA(V)	1	0.21	GCA(A)	2	0.53	GAA(E)	2	1.33	GGA(G)	6	1.41
GUG(V)	13	2.74	GCG(A)	4	1.07	GAG(E)	1	0.67	GGG(G)	2	0.47

FAdV-5/B

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	2	1	UCU(S)	0	0	UAU(Y)	4	0.8	UGU(C)	0	0
UUC(F)	2	1	UCC(S)	3	1.13	UAC(Y)	6	1.2	UGC(C)	1	2
UUA(L)	1	0.46	UCA(S)	0	0	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	3	1.38	UCG(S)	4	1.5	UAG(*)	0	0	UGG(W)	1	1
CUU(L)	3	1.38	CCU(P)	6	1.85	CAU(H)	1	2	CGU(R)	2	4
CUC(L)	1	0.46	CCC(P)	3	0.92	CAC(H)	0	0	CGC(R)	0	0
CUA(L)	0	0	CCA(P)	0	0	CAA(Q)	2	1	CGA(R)	0	0
CUG(L)	5	2.31	CCG(P)	4	1.23	CAG(Q)	2	1	CGG(R)	0	0
AUU(I)	2	1	ACU(T)	1	0.19	AAU(N)	4	0.42	AGU(S)	3	1.13
AUC(I)	3	1.5	ACC(T)	10	1.9	AAC(N)	15	1.58	AGC(S)	6	2.25
AUA(I)	1	0.5	ACA(T)	0	0	AAA(K)	2	0.8	AGA(R)	0	0
AUG(M)	6	1	ACG(T)	10	1.9	AAG(K)	3	1.2	AGG(R)	1	2
GUU(V)	1	0.44	GCU(A)	6	1.09	GAU(D)	3	1	GGU(G)	2	0.67
GUC(V)	2	0.89	GCC(A)	11	2	GAC(D)	3	1	GGC(G)	2	0.67
GUA(V)	4	1.78	GCA(A)	0	0	GAA(E)	3	1.2	GGA(G)	8	2.67
GUG(V)	2	0.89	GCG(A)	5	0.91	GAG(E)	2	0.8	GGG(G)	0	0

FAdV-2/11/D

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	3	1.2	UCU(S)	1	0.55	UAU(Y)	4	0.73	UGU(C)	0	0
UUC(F)	2	0.8	UCC(S)	0	0	UAC(Y)	7	1.27	UGC(C)	0	0
UUA(L)	1	0.6	UCA(S)	2	1.09	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	1	0.6	UCG(S)	4	2.18	UAG(*)	0	0	UGG(W)	2	1
CUU(L)	2	1.2	CCU(P)	2	0.62	CAU(H)	0	0	CGU(R)	0	0
CUC(L)	4	2.4	CCC(P)	7	2.15	CAC(H)	0	0	CGC(R)	1	1.5
CUA(L)	2	1.2	CCA(P)	0	0	CAA(Q)	2	0.8	CGA(R)	2	3
CUG(L)	1	0.6	CCG(P)	4	1.23	CAG(Q)	3	1.2	CGG(R)	0	0
AUU(I)	2	0.86	ACU(T)	3	0.75	AAU(N)	5	0.77	AGU(S)	0	0
AUC(I)	5	2.14	ACC(T)	4	1	AAC(N)	8	1.23	AGC(S)	4	2.18
AUA(I)	0	0	ACA(T)	4	1	AAA(K)	4	1.14	AGA(R)	0	0
AUG(M)	5	1	ACG(T)	5	1.25	AAG(K)	3	0.86	AGG(R)	1	1.5
GUU(V)	2	0.62	GCU(A)	5	1	GAU(D)	4	0.73	GGU(G)	3	0.86
GUC(V)	9	2.77	GCC(A)	10	2	GAC(D)	7	1.27	GGC(G)	5	1.43
GUA(V)	0	0	GCA(A)	2	0.4	GAA(E)	5	1.67	GGA(G)	4	1.14
GUG(V)	2	0.62	GCG(A)	3	0.6	GAG(E)	1	0.33	GGG(G)	2	0.57

FAdV-8a/E

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	2	0.8	UCU(S)	3	1.2	UAU(Y)	2	0.33	UGU(C)	0	0
UUC(F)	3	1.2	UCC(S)	2	0.8	UAC(Y)	10	1.67	UGC(C)	0	0
UUA(L)	0	0	UCA(S)	0	0	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	1	0.55	UCG(S)	1	0.4	UAG(*)	0	0	UGG(W)	1	1
CUU(L)	0	0	CCU(P)	5	1.54	CAU(H)	0	0	CGU(R)	1	1.5
CUC(L)	2	1.09	CCC(P)	6	1.85	CAC(H)	0	0	CGC(R)	2	3
CUA(L)	2	1.09	CCA(P)	0	0	CAA(Q)	3	1.2	CGA(R)	0	0
CUG(L)	6	3.27	CCG(P)	2	0.62	CAG(Q)	2	0.8	CGG(R)	0	0
AUU(I)	1	0.5	ACU(T)	2	0.47	AAU(N)	4	0.62	AGU(S)	1	0.4
AUC(I)	5	2.5	ACC(T)	5	1.18	AAC(N)	9	1.38	AGC(S)	8	3.2
AUA(I)	0	0	ACA(T)	3	0.71	AAA(K)	2	0.8	AGA(R)	1	1.5
AUG(M)	5	1	ACG(T)	7	1.65	AAG(K)	3	1.2	AGG(R)	0	0
GUU(V)	0	0	GCU(A)	3	0.67	GAU(D)	1	0.33	GGU(G)	5	1.11
GUC(V)	7	2.33	GCC(A)	8	1.78	GAC(D)	5	1.67	GGC(G)	5	1.11
GUA(V)	2	0.67	GCA(A)	1	0.22	GAA(E)	4	1	GGA(G)	2	0.44
GUG(V)	3	1	GCG(A)	6	1.33	GAG(E)	4	1	GGG(G)	6	1.33

FAdV-4/C

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	1	0.67	UCU(S)	0	0	UAU(Y)	2	0.31	UGU(C)	0	0
UUC(F)	2	1.33	UCC(S)	7	3.5	UAC(Y)	11	1.69	UGC(C)	1	2
UUA(L)	0	0	UCA(S)	1	0.5	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	0	0	UCG(S)	2	1	UAG(*)	0	0	UGG(W)	2	1
CUU(L)	0	0	CCU(P)	2	0.57	CAU(H)	0	0	CGU(R)	1	1.5
CUC(L)	4	3.43	CCC(P)	9	2.57	CAC(H)	0	0	CGC(R)	1	1.5
CUA(L)	0	0	CCA(P)	1	0.29	CAA(Q)	0	0	CGA(R)	0	0
CUG(L)	3	2.57	CCG(P)	2	0.57	CAG(Q)	5	2	CGG(R)	0	0
AUU(I)	1	0.5	ACU(T)	0	0	AAU(N)	4	0.47	AGU(S)	0	0
AUC(I)	3	1.5	ACC(T)	9	2.12	AAC(N)	13	1.53	AGC(S)	2	1
AUA(I)	2	1	ACA(T)	0	0	AAA(K)	1	0.4	AGA(R)	1	1.5
AUG(M)	3	1	ACG(T)	8	1.88	AAG(K)	4	1.6	AGG(R)	1	1.5
GUU(V)	0	0	GCU(A)	3	0.71	GAU(D)	3	0.86	GGU(G)	5	1.33
GUC(V)	5	1.82	GCC(A)	7	1.65	GAC(D)	4	1.14	GGC(G)	5	1.33

FAdV-8b/E

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	3	1.2	UCU(S)	3	1.5	UAU(Y)	2	0.33	UGU(C)	0	0
UUC(F)	2	0.8	UCC(S)	0	0	UAC(Y)	10	1.67	UGC(C)	0	0
UUA(L)	2	1	UCA(S)	2	1	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	1	0.5	UCG(S)	1	0.5	UAG(*)	0	0	UGG(W)	1	1
CUU(L)	2	1	CCU(P)	7	2.15	CAU(H)	0	0	CGU(R)	0	0
CUC(L)	1	0.5	CCC(P)	3	0.92	CAC(H)	0	0	CGC(R)	1	1.5
CUA(L)	1	0.5	CCA(P)	1	0.31	CAA(Q)	4	1	CGA(R)	0	0
CUG(L)	5	2.5	CCG(P)	2	0.62	CAG(Q)	4	1	CGG(R)	2	3
AUU(I)	1	0.5	ACU(T)	1	0.24	AAU(N)	1	0.15	AGU(S)	2	1
AUC(I)	5	2.5	ACC(T)	6	1.41	AAC(N)	12	1.85	AGC(S)	4	2
AUA(I)	0	0	ACA(T)	5	1.18	AAA(K)	3	1	AGA(R)	1	1.5
AUG(M)	4	1	ACG(T)	5	1.18	AAG(K)	3	1	AGG(R)	0	0
GUU(V)	2	0.8	GCU(A)	4	0.84	GAU(D)	1	0.29	GGU(G)	2	0.47
GUC(V)	2	0.8	GCC(A)	8	1.68	GAC(D)	6	1.71	GGC(G)	5	1.18

Everage value

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	1.9	0.87	UCU(S)	1.4	0.66	UAU(Y)	2.6	0.44	UGU(C)	0	0
UUC(F)	2.4	1.13	UCC(S)	2.3	1.05	UAC(Y)	9	1.56	UGC(C)	0.4	2
UUA(L)	0.6	0.32	UCA(S)	1.3	0.59	UAA(*)	0	0	UGA(*)	0	0
UUG(L)	1	0.55	UCG(S)	2	0.92	UAG(*)	0	0	UGG(W)	1.4	1
CUU(L)	1.3	0.71	CCU(P)	4.1	1.26	CAU(H)	0.1	2	CGU(R)	0.9	1.2
CUC(L)	2.6	1.42	CCC(P)	5.6	1.7	CAC(H)	0	0	CGC(R)	1	1.4
CUA(L)	0.9	0.47	CCA(P)	0.6	0.17	CAA(Q)	2.3	0.73	CGA(R)	0.4	0.6
CUG(L)	4.6	2.53	CCG(P)	2.9	0.87	CAG(Q)	4	1.27	CGG(R)	0.4	0.6
AUU(I)	1.6	0.77	ACU(T)	1.4	0.33	AAU(N)	3.6	0.47	AGU(S)	1.6	0.73
AUC(I)	4	1.95	ACC(T)	6.9	1.6	AAC(N)	11.5	1.53	AGC(S)	4.4	2.04
AUA(I)	0.6	0.28	ACA(T)	2.6	0.6	AAA(K)	1.9	0.7	AGA(R)	0.9	1.2
AUG(M)	4.3	1	ACG(T)	6.3	1.47	AAG(K)	3.4	1.3	AGG(R)	0.7	1
GUU(V)	1.3	0.43	GCU(A)	4	0.88	GAU(D)	2.1	0.59	GGU(G)	3.1	0.82
GUC(V)	4.3	1.43	GCC(A)	8	1.76	GAC(D)	5.1	1.41	GGC(G)	4.7	1.23
GUA(V)	1.6	0.52	GCA(A)	2.3	0.5	GAA(E)	3.1	1.02	GGA(G)	4.3	1.12

Fig. 5 (A-G) RSCU for FAdV-A-E. Analysis of the number of successive codons and relative synonymous codon usage. Synonymous codons are indicated in colours. The colour intensity indicates how highly an amino acid is preferred in a particular position amongst species. Deeper colours are used for the preferable codons.

preferred for codes for Leu (L) (RSCU accordingly 2.4 and 3.43). In the case of FAdV-4/C, only codons CUC and CUG were present out of six possibilities. Similarly, FAdV-1/A, FAdV-4/C, FadV8a/E, and strains from FAdV-2/11/D prefer codon CCC which codes for Pro (P); however, in the case of FAdV-5/B, FAdV-7/E, or FAdV-8b/E the most frequent codon was CCU (Fig 5. A-G).

The codon preference is a sign of natural selection, and is present in the whole genome sequence, however, its intensity was lower than in the Loop L1 region (Niczyporuk, 2017). The average values of RSCU in the genome of the FAdV-1A reference strain was similar to the average values of RSCU in the Loop L1 HVRs-1-4

of the examined strains. For example, the RSCU value was 2.53 for the codon CUG codes for Leu (L), whereas this value was 1.37 for the genome of the FAdV-1/A reference strain. Values for the remaining codon codes for Leu (L) were higher indicating more regular frequency of codon appearance. In the case of codon codes for Arg (A) in Loop L1, codon CGC, and CGG appeared with the RSCU values 1.4 and 0.6 respectively, but in the whole genome sequence for FAdV-1/A these values were accordingly 1.13 and 1.16. Similarly, in the case of codon codes for Ala (A) in Loop L1, codon GCC was preferable with RSCU 1.76, whereas in the FAdV-1/A genome it was codon GCG (RSCU 1.22) (Fig. 5 A-G).

Analysis of nucleotide codon composition in the examined region for type/species FAdV-1/A, and FAdV-7/E was presented by Niczyporuk, 2017. In this study the RSCU obtained for the other types/species isolated in Poland is presented in Fig. 5 (A-G).

A very interesting interdependence can be seen in the case of nine codon preferences. The codons in which G or C can appear on the 3rd site are preferable in HVRs-1-4. Analysis of nucleotide composition and codon adequacy proved this dependency. In the case of consensus sequence in examined types/species in 68.7% of codons on the 3rd position were G (26.8%) or C (41.9%). For separate FAdV species this value ranged from 63.3% in the case of FAdV-2/11/D to 79% in the case of FAdV-4/C; the values for GC for separate species ranged from 54.2% for FAdV-2/11/D and FAdV-5/B, to 59.7% for FAdV-4/C. This indicates the higher percentage of GC presence on the 3rd position in the codon than in other positions, which has a great impact on codon composition and amino acid coding.

Discussion

The degeneracy of the genetic code is establish with a group of 2-6 synonymous codons, and some codons can be preferred to others. This specific use of codons is called codon usage bias. Among the causative various of factors, natural selection, mutational pressure, RNA structure, and gene length are most important for codon usage bias (Singh and Tyagi 2017). Codon usage analysis is a well-established technique for understanding the process of evolution at the molecular level (Choludhury et al. 2017). In our study the nucleotide frequency in codons was determined by GC pairs content, and was different for different types/species. The value obtained was between 54.2% for FAdV-2/11/D and FAdV-5B, to 59.7% for FAdV-4/C, with an average of about 56.3%. The data obtained are comparable to those reported by Raue et al. (2005). It may be possible that genes of higher transcriptional importance could have a higher GC content (Epstein et al. 2000, Hassan et al. 2009). Similar research concerning different species has been done by Halder et al. (2017). Every deviation in codon usage is based on the codon preference by using the exact codon during the translation process. A specific codon is used more frequently than the other synonymous codon in genes under high expression. Such a codon is referred to as prefereble or optimal. This adaptation is to force expression, gene size, genome structure, and the percentage of GC content or the frequency of recombinations. In this study an analysis of codon usage for FAdV-2/11, 5, 8a, and 8b was

performed to establish and evaluated if, in the HVRs-1-4 region of the hexon gene, the preference codon appeared or not. RSCU for FAdV-1/A and 7/E was reported by Niczyporuk (2017), and preference of codons was indicated. Fig. 6 (A-G) shows the analysis of the number of codons and the relative synonymous codon usage.

The next step, the analysis of the amino acid sequence, was performed. The Loop L1 with HVRs-1-4 is the main indicator of variability, as confirmed by other researchers (Crawford-Miksza and Schnurr 1996, Raue et al. 2005, Pichla-Gollon et al. 2007, Yu et al. 2012). Amino acid sequences were obtained after the "virtual" translation of consensus nucleotide sequences for each examined adenovirus type/species. The sequences of 176aa long were created and analysed. Alignment analysis confirmed the locations of the HVR1-4 regions with 122 amino acids positions. Variations in amino acids in four HVR1-4 were determined. The analysis indicated that the amino acid HVR1-4 sequence of FAdV-4/C is the most divergent from the examined consensus sequences. Not every amino acid substitution has an influence on structure and protein function. Most of the substitutions are of similar size, charge, or hydrophobic property. These are the amino acids with conservation function. This fact can explain that a protein may have self-structure and function with different amino acid sequences coded by highly different nucleotide sequences. Amino acid analysis in Loop L1 indicated, that only 61 amino acid positions (36.2%) of 177 examined were identical in all examined strains (consensus sequences), but in HVR1-4 regions, only 29 amino acid positions (23.8%) were conservative.

We can establish, that every codon can appear with equal frequency; however in this study indicated that, from several possibilities, only one codon can be preferable. This preferences can appear in genes with strong expression (Hassan et al. 2009, Behura and Severson 2012). The frequency of the presence of some codons in front of others (codon synonymity) is called RSCU. The optimal codon can lead to fast and exact or accurate translation. This is extremely important for proteins which are synthesised in greater amounts (Xu et al. 2007). This means that they have an impact in evolution. Analysis of examined consensus sequences based on the evaluation of mutation quantity, their localization and possibility of the influence on protein tertiary structure (data not yet published) has been performed. Many authors suggest (Crawford-Miksza and Schnurr 1996, Epstein et al. 2000, Behura and Severson 2012) that the mutations which are located on the 1st and 2nd codon positions are the most important for study. These are mutations which have an influence on amino acid coding influencing differences in protein

structure and function. Changes in variability have, in consequence, an impact on virus pathogenicity. Principal component analysis also supported the view, that most codons showed a biased effect on G and C at the 3rd codon position and preferred codons which ended with either G or C. Halder et al. (2017) also indicated that a positive significant correlation in gene expression parameter with a few amino acids such as Val (V), Arg (A), Ser (S), and Ile (I) might influence gene expression.

The analysis of nucleotide sequence indicated, that different codons can code this same amino acid, however, some of them are preferable. Codon CUC existing in strains: FAdV-4/C or FAdV-2/11/D codes for lysine Lys (L); however this function was used by the CUG codon on the rest of the examined strains. The study indicated that the usage of percentage GC pairs for separate types/species was between 54.2% for FAdV-2/11/D and FAdV-5/B, and 59.7% for FAdV-4/C, with an average value of 56.3%. The data obtained are similar to those presented by Raue et al. 2005. We can assume that genes of higher transcription importance value have a higher percentage of GC pair content.

Gu et al. (2012) compared the RSCU of ovine 287(OAdV287) and human HAdV2/5 adenovirus strains. They found that OAdV287 had more conservative codon usage than HAdV 2/5. The preference codons of HAdV2/5 mostly had GC ends and this was a similar result compared with our own study. Das et al. (2012) indicated that one of the major determinants is the GC content in the 3rd codon position (C3S and G3S), and significant variations are observed in synonymous codon choice in structural and nonstructural genes of HAdV. A previous study by Niczyporuk, 2017 described the RSCU in the Loop L1 region of the hexon gene for types/species FAdV-1/A, and FAdV-7/E. The most important mutations were those in the 1st and 2nd codon positions, since these mutations are more likely to result in an amino acid change affecting the structure and protein function. Nucleotide sequence analysis indicated that different codons can code the same amino acid, but some of them are preferred. Codon analysis of the Loop L1 region of the hexon gene indicated differences in codon preference patterns between adenovirus strains representing diverse types. It was found that C was the most frequent nucleotide for each type ranging from RSCU 29.3 to RSCU 34.4.

RSCU analysis revealed that AGA and CTG codons were over-represented in Y-linked genes (Choludhury et al. 2017). Analysis of the average number of codons indicated that the variation in codon usage is presented in every FAdV species. Similar RSCU analysis was carried out by Singh and Tyagi

(2017) for ZIKV, in humans and mosquitoes.

In most cases in codon preferences in adenovirus genome are closely similar which we can observe in the case of every values of examined types/species. As the example of the CUG codon coding for Leu (L) with RSCU 2.53 shows, compared to whole genome FAdV-A/1 with (RSCU 1.37), its value for the rest of Leu (L) codons is higher, as indicated in this study.

In conclusion, codon usage analysis of the FAdVs HVR-1-4 region of the hexon gene can be further supportive for developing novel diagnostic methods or inheriting the evolutionary origin of HVR-1-4 with other AdVs strains. Such analysis provide information regarding the high or low expression genes which can be selected. The study indicated that the greater the distance between types/species, the greater is the variation in codon usage. Analysis indicated that variations in codon usage for FAdVs are types/species specific and it is determined by random mutation over the evolutionary process. The present work is the codon analysis of a FAdV Loop L1 region of the hexon gene, and this is the first study conducted on RSCU on the FAdVs 2/11/D, 4/C, 5/B, 8a/E, and 8b/E in Poland.

Conclusion

The analysis indicated antigenic properties of the examined viruses, the presence of relative synonymous codon usage, and the presence of mutations. The effect of mutational pressure on the codon can be tested in future in order to understand its impact on FAdV pathogenicity.

References

Behura SK, Severson DW **(2012)** Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PLoS One 7: e43111

Beladi I, Pusztai R, Bakay M, Mucsi I, Tarodi B **(1979)** Interferon induction by adenoviruses. Postgrad Med J 55: 128-134.

Benko M, Harrach B, Both G, Russel W, Adair B, Adam E, Jong J, Hess M, Johnson M, Kajon A, Kidd AH, Lehmkuhl HD, Li QG, Mautner V, Pring-Akerblom P, Wadell G **(2005)** Family Adenoviridae. 1 st ed., In: Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L Virus Taxonomy. Eighth Report of the International Committee on the Taxonomy of Viruses. Elsevier, New York, p. 213-228.

Berencsi K, Rando RF, de Taisne C, Paoletti E, Plotkin SA, Gonczol E **(1993)** Murine cytotoxic T cell response specific for human cytomegalovirus glycoprotein B (gB) induced by adenovirus and vaccinia virus recombinants expressing gB. J Gen Virol 74 : 2507-2512.

Chiocca S, Kurzbauer R, Schaffner G, Baker A, Mautner V, Cotten M **(1996)** The complete DNA sequence and genomic organization of the avian adenovirus CELO. J Virol 70: 2939-2949.

Choludhury M, Uddin A, Chakraborty S **(2017)** Codon usage bias and its influencing factors for Y-linked genes in human. Comput Biol Chem 69: 77-86.

Corredor JC, Garceac A, Krell PJ, Nagy E **(2008)** Sequence comparison of the right end of fowl adenovirus genomes. Virus Genes 36: 331-344.

Crawford-Miksza L, Schnurr DP **(1996)** Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype--specific residues. J Virol 70: 1836-1844.

Das S, Paul S, Dutta C **(2012)** Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res 117: 227-236.

Epstein RJ, Lin K, Tan TW **(2000)** A functional significance for codon third bases. Gene 245: 291-298.

Fox JM, Erill I **(2010)** Relative codon adaptation: a generic codon bias index for prediction of gene expression. DNA Res 17: 185-196.

Gu YX, Zhang J, Zhou JH, Zhao F, Liu WQ, Wang M, Chen HT, Ma LN, Ding YZ, Liu YS **(2012)** Comparative analysis of ovine adenovirus 287 and human adenovirus 2 and 5 based on their codon usage. DNA Cell Biol 31: 360-366.

Halder B, Malakar AK, Chakraboty S **(2017)** Nucleotide composition determines the role of translational efficiency in human genes. Bioinformation 13: 46-53.

Harrach B **(2008)** Adenoviruses: General Features, p. 1-9. Mahy BWJ, RMHV 1 st ed., Encyclopedia of Virology (Third Edition). Academic Press of Elsevier, Oxford, UK

Harrach B, Kajan G **(2011)** Aviadenovirus. 1 st ed. In: Darai G, Konle S, Simniok J Aviadenoviridae Springer, Berlin, p. 13-28.

Hassan S, Mahalingam V, Kumar V **(2009)** Synonymous codon usage analysis of thirty two mycobacteriophage genomes. Adv Bioinformatics 2009: 316936.

Kay MA, Holterman AX, Meuse L, Gown A, Ochs HD, Linsley PS, Wilson CB **(1995)** Long-term hepatic adenovirus-mediated gene expression in mice following CTLA4Ig administration. Nat Genet 11: 191-197.

Knowles DP **(2011)** Adenoviridae. 1 st ed. In: MacLachlan NJ, Dubovi EJ Fenner's Veterinary Virology, ed., Academc Press of Elsevier, Oxford, UK, p. 203-212.

Niczyporuk JS **(2014)** Molecular characteristic on occurrence of fowl adenovirus field strains and effect of efficiency on prophylactic vaccinations against Marek's disease. Doctoral dissertation, National Veterinary Research Institute, Pulawy, Poland

Niczyporuk JS **(2017)** Molecular characterisation of fowl adenovirus type 7 isolated from poultry associated with inclusion body hepatitis in Poland. Arch Virol 162: 1325-1333.

Pichla-Gollon SL, Drinker M, Zhou X, Xue F, Rux JJ, Gao GP, Wilson JM, Ertl HC, Burnett RM, Bergelson JM **(2007)** Structure-based identification of a major neutralizing site in an adenovirus hexon. J Virol 81: 1680-1689.

Raue R, Gerlach H, Muller H **(2005)** Phylogenetic analysis of the hexon loop 1 region of an adenovirus from psittacine birds supports the existence of a new psittacine adenovirus (PsAdV). Arch Virol 150: 1933-1943.

Russell WC **(2000)** Update on adenovirus and its vectors. J Gen Virol 81: 2573-2604.

Rux JJ, Kuser PR, Burnett RM **(2003)** Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol 77: 9553-9566.

Singh NK, Tyagi A **(2017)** A detailed analysis of codon usage patterns and influencing factors in Zika virus. Arch Virol 162: 1963-1973.

Sharp PM, Li WH **(1986)** An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24: 28-38.

Toth MI, Arya B, Pusztai R, Shiroki K, Beladi I **(1987)** Interferon induction by adenovirus type 12: stimulatory function of early region 1A. J Virol 61: 2326-2330.

Xu L, Benson SD, Burnett RM **(2007)** Nanoporous crystals of chicken embryo lethal orphan (CELO) adenovirus major coat protein, hexon. J Struct Biol 157: 424-431.

Yu B, Wang C, Dong J, Zhang M, Zhang H, Wu J, Wu Y, Kong W, Yu X **(2012)** Chimeric hexon HVRs protein reflects partial function of adenovirus. Biochem Biophys Res Commun 421: 170-181.